

STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 151245

TO: Ginny Portner
Location: REM-3B02&3C4C70
Art Unit: 1645
Monday, April 25, 2005

Case Serial Number: 10/039183

From: Deirdre Arnold
Location: Biotech-Chem Library
REM 1A64
Phone: 571-272-2532

Deirdre.Arnold@uspto.gov

Search Notes

Please feel free to contact me if you have any questions or would like to amend the search.

Thank you for using STIC services.

Regards,
Deirdre Arnold



HPD

GBD1360 SED ID

QY 181 Prolysalalysylglu1alysPhe1leu1AAsnArgAspThr1leAAspPro 200
 DB 646 CCAAGGCTRAAAAAGACCTTAATTCATTGACTTGCCATTCGGATTCGATTCCT 705
 QY 201 AAsnSer1yAsnAlaGlnaAng1yG1yAAspLeuG1yLysPheGln1yAsnGln1yMetAla 220
 DB 706 AACACAGAGAACCGCGCAAAATGCGGTGATTGGGGAATTCGAAAGAACCAATGGCT 765
 QY 221 ProAAspPheSer1yLysAla1AAspHe1AeuthrProG1yAAspTyThr1yThrProVal 240
 DB 766 CCGGATTTTCTTAAGCCCGCTTCCTTACCTCCGCGGATTAACCTAAACCCCTGTT 825
 QY 241 LysThrGluPheG1yTyH1s1le1yTyLeu1leSer1yAspSerProVal1ThrTy 260
 DB 826 AAAACAGAGATTGGTTATCATATTATCTATTGATTTCTTAAGATAGCCCTGTACTTAT 885
 QY 261 ThrTyGluGln1Ala1yAspProThr1le1yG1yMetLeuGln1yLysLeuPheGln1y 280
 DB 886 ACTTATGAACAGGCTTAACCTTAACCTTAAGGGGATGTTCAAGAAAGCTTTTCCAGAA 945
 QY 281 ArgMetAAsnGlnArg1leGluGluLeuArg1yH1sAla1yLys1leVal1leAsn1y 299
 DB 946 CCCATGAATCAACGCGATTGAGGAACTAAGAAAGCAGCTAAATGTTTATCAACAAG 1002

RESULT 3
 LOCUS BD082346 1149 bp DNA linear PAT 27-AUG-2002

DEFINITION 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and corresponding polynucleotide molecules.

ACCESSION BD082346
 VERSION BD082346.1 GI:22627956
 KEYWORDS JP 2001523954-A/53.

SOURCE Mabsadenovirus
 ORGANISM Mabsadenovirus
 1 (bases 1 to 1149)
 Viruses; dsDNA viruses, no RNA stage; Adenoviridae.

REFERENCE Klenzhausen, H., Lissolo, L., Tomb, J. F., Miller, C. and Garawi, A. A.
 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and corresponding polynucleotide molecules
 Patent: JP 2001523954-A 53 27-NOV-2001;
 MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF; PASTEUR MERIEUX SERUMS ET VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
 HUMAN GENOME SCIENCES INC

COMMENT
 PN JP 2001523954-A/53
 PD 27-NOV-2001
 PF 31-MAR-1998 JP 1998541962

PI HAROLD KLEANTHOSU, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI MILLER,
 PI AMAL AL GARAWI
 PC A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,
 PC A61K39/40,
 PC G01N33/554, G01N33/569
 CC Serendehness: Single;
 CC Topology: linear;
 FH key Location/Qualifiers.

FEATURES
 Source Location/Qualifiers
 1. 1149
 /organism="Mabsadenovirus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10509"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,31e-108 Length: 1149
 Score: 1520.00 Matches: 299
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-039-183a-2 (1-299) x BD082346 (1-1149)

QY 1 MetLysLysAsn1leuAAsn1leuVal1G1yAla1euserTherSerPheLeuMet 20
 DB 106 ATGAAAAAATATCTTAATTTAGCGTTAGGGGCGCTTGAGCAGCGTGGTTTGATG 165
 QY 21 AlaLysProAlaH1sAsn1AAsnAsn1eThH1sAsnThLys1yThr1yAspSer 40
 DB 166 GCTAAGCCCGCTCATACCAATATACCTAGCAGCATTAACGAAAAAAGCACTGATTC 225
 QY 41 SerAlaG1yVal1leuAlaTherVal1eSp1yAAspPro1leThLysSer1yAspPhe 60
 DB 226 TCAGCAGGCGGTGTAGCGACGTGATGGCAGACCTATACCTAAAGCAGATTTTGACATG 285
 QY 61 IleLysGlnArgAspProAAspPheAspPheAspLysLeu1yG1yLysG1yLysAla 80
 DB 286 ATTAAGCAACGAAATCTTAATTTGATTTGACAGCTTAAGAGAGAAAGAAAGAACCC 345
 QY 81 Ileu1eAspGln1Ala1leArgThra1leuVal1G1yAAsnGln1Ala1yThrG1yLys 100
 DB 346 TTGATTTGATCAAGCTATTCGACACCGCTTGTAGAAATGAAAGCTAAACCGAGAAATGG 405
 QY 101 AAspSerThrProGluPheLysAlaMetMetG1yAlaVal1yLysGln1Ala1eVal1Glu 120
 DB 406 GACAGACCTCCAGAAATTTAAAGCATGATGAGAGCGGTTAAAGACAGGCTTTAGTGGA 465
 QY 121 PheTTPAlaLysLysGln1AgluGluVal1yLysVal1Gln1leProGluLysGluMet 140
 DB 466 TTTTGGGCTTAAGAAACAGGCTGAGAGAGTGAAGAAAGTCCAAATCCAGAAAGAAAGATG 525
 QY 141 GlnAAspPheTyAsnAla1eAsn1yAspGln1eAspVal1yLysGln1Ala1eH1sAlaArg 160
 DB 526 CAAGATTTTTCACAGCTTAACAAAGATCAGCTTTTGTTCACAGCAAGAACCCATGCTAG 585
 QY 161 H1s1leuVal1yLysThrGluAspGluAla1yAsnArg1le1leSerGlu1leAsp1yGln 180
 DB 586 CATATTTTATGTAAGAAACCGAAGATGAGCTTAACCGATTAATTTCTGAGATTGACAAACAG 645
 QY 181 Prolysalalysylglu1alysPhe1leu1AAsnArgAspThr1leAAspPro 200
 DB 646 CCAAGGCTRAAAAAGACCTTAATTCATTGACTTGCCATTCGGATTCGATTCCT 705
 QY 201 AAsnSer1yAsnAlaGlnaAng1yG1yAAspLeuG1yLysPheGln1yAsnGln1yMetAla 220
 DB 706 AACACAGAGAACCGCGCAAAATGCGGTGATTGGGGAATTCGAAAGAACCAATGGCT 765
 QY 221 ProAAspPheSer1yLysAla1AAspHe1AeuthrProG1yAAspTyThr1yThrProVal 240
 DB 766 CCGGATTTTCTTAAGCCCGCTTCCTTACCTCCGCGGATTAACCTAAACCCCTGTT 825
 QY 241 LysThrGluPheG1yTyH1s1le1yTyLeu1leSer1yAspSerProVal1ThrTy 260
 DB 826 AAAACAGAGATTGGTTATCATATTATCTATTGATTTCTTAAGATAGCCCTGTACTTAT 885
 QY 261 ThrTyGluGln1Ala1yAspProThr1le1yG1yMetLeuGln1yLysLeuPheGln1y 280
 DB 886 ACTTATGAACAGGCTTAACCTTAACCTTAAGGGGATGTTCAAGAAAGCTTTTCCAGAA 945

RESULT 4
 LOCUS AE000538 11421 bp DNA linear BCT 06-APR-1999
 DEFINITION Helicobacter pylori 26695 section 16 of 134 of the complete genome.
 ACCESSION AE000538 AE000511
 VERSION AE000538.1 GI:2313263
 KEYWORDS
 SOURCE
 ORGANISM
 Helicobacter pylori 26695
 Helicobacter pylori 26695
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 Helicobacteraceae; Helicobacter.
 1 (bases 1 to 11421)
 Tomb, J.-F., White, O., Kellavage, A. R., Clayton, R. A., Sutton, G. G.,

```

TITLE
JOURNAL Nature 388 (6642), 539-547 (1997)
MEDLINE 97394467
PUBMED 9252185
AUTHORS
2 (bases 1 to 11421)
Tom, J., F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G.,
Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S.,
Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F.,
Glodek, A., McKenney, K., Fitzgerald, L. M., Lee, N., Adams, M. D.,
Hickey, E. K., Berg, D. E., Gocayne, J. D., Uterback, T. R.,
Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C.,
Bowman, C., Matthews, L., Wallin, E., Hayes, W. S., Borodovsky, M.,
Karp, P. D., Smith, H. O., Fraser, C. M. and Venter, J. C.
The complete genome sequence of the gastric pathogen Helicobacter
pylori
SOURCE
/organism="Helicobacter pylori 26695"
/mol_type="genomic DNA"
/strain="26695"
/db_xref="taxon:85962"
246.1145
/gene="HP0175"
246.1145
/gene="HP0175"
246.1145
/gene="HP0175"
/feature="similar to GB:671840 SP:Q46105 percent identity:
34.90; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell binding factor 2"
/protein_id="AAD07245.1"
/db_xref="GI:2313264"
/transl_table=11
/translation="MKNTLNLALVGAISTFLMAKPAHANNATNTKTTDSAGV
LATVGRPTTSDPMIKQNPDPFDKKEKEKALIDQARTALVENEATKELDS
TPEFAMAEAVKQALVEFWAKQAEVQVQIPEKMQDPFNANKDQLFVQEAHAR
HILVTEDEAKRIISIDIKQPAKKEAKPIELANDDTIDPNSKNONGDGLKPOKNO
MAPDSKAPALTPGDYTKTPVTERGHIITLISKDSPTVYTYEQAKPTIKGMQOK
LFOERNQRIEELRKAIVINK"
1162.2085
/gene="HP0176"
1162.2085
/gene="HP0176"
1162.2085
/gene="HP0176"
/feature="similar to GB:M22039 SP:P13243 PID:460911
PID:853785 GB:M009126 percent identity: 45.96; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="fructose-bisphosphate aldolase (tet)"
/protein_id="AAD07246.1"
/db_xref="GI:2313265"
/transl_table=11
/translation="MLVKGNEILLRAHKEGVGVAFNPNFEMLNAIPEAGENSPL
FIQSEGAIKWGDIMAVGMVTCRCRYPHIVLALHDTGTFESCEKAVKGFVSM
IDASHAEENELTSTKVYKMAHAGVSEALVLEKIEDNI SVDEKAVNAGFSA
EOPVKEQVDYLAIPALIGTSHGAFKFKGKLEDFEELQEKRLTNIPLVHGSAIPDN
VRKSYLDAGDGLGSGKVPFELQESVGVGINKWNTDIDRLAIFAEVRKANDEKSO
FDLRKFSFPAQLAKNVVKKRMLGSAKNT"
2107.2670
/gene="HP0177"
/feature="similar to GB:M22039 SP:P13243 PID:460911
PID:536991 percent identity: 45.05; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="translation elongation factor EF-P (efp)"
/protein_id="AAD07247.1"
/db_xref="GI:2313266"
/transl_table=11
/translation="MAIGMSLEKGLKXIEIGVPRYIYGVHVRKGAAPYAKIKS
FLDGKVIKTHAGDKCEPNIVETQMOTYHDTGTPQPMDSKQALNDQSGEA
SKMMDGMQVQVLLHNDRAISVDVQVALKIVETAPNPKGDTSSASKPAVLEGA
VQVPRVHVEGEIIRKIVTETELKVK"
complement(4205)
/gene="HP0178"
complement(3179)
/gene="HP0178"
complement(4205)
/feature="similar to GB:L77117 PID:1591717 percent identity:
36.25; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="spore coat polysaccharide biosynthesis protein
E"
/protein_id="AAD07248.1"
/db_xref="GI:2313267"
/transl_table=11
/translation="MLQPKTIVAEISANHNDLNLAKESLHAIKESGDFVQLQYTP
SCMTLNSKEDPTIGTGLMDKENVLYOKASTPLEMHAELRLKIDLIQSFSPS
SOALEBELSINCPMYKIASPFIYDILBKARTQPKIISGIAITHEDQALICR
RVNNDITLLKCVSAVPSKIEDANLISWVKGEIGVFGSDHTIGSLCPLATTLG
ASMIKHFILINKSLQTPDSARFSDPNFGSKVNEALQSVLALGEEEPINPTLTKR
FFARSLFVIRKDIQKEALTENNIKALRENLGHPKFEYKEIIGOKASKELKANTPLSAD
DIERSL"
complement(4205)
/gene="HP0179"
complement(4205)
/gene="HP0179"
complement(4205)
/feature="similar to GP:1786703 percent identity: 37.62;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAD07249.1"
/db_xref="GI:2313268"
/transl_table=11
/translation="MIRALINISHAPEKPLVNGVNLHIREKESAILIGVSGSKSTLLS
HLATMLKNSGTSILSHODIYALNSKLELRLKVGIIPOSHVLPGFSALENDVQ
ASILAKQKINSLEQLQIATITLKGVEELSGGQOQRASITARVSKKRIITADEPTG
NLDVTSANQVISMQLQNYITEKGAVALTHDEHLAFTCSQVTRKEVLIKSK"
complement(4843)
/gene="HP0180"
complement(4843)
/gene="HP0180"
complement(4843)
/feature="similar to GB:L42023 SP:P4626 PID:1003497
PID:1222229 PID:1204555 percent identity: 28.03;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="apolipoprotein N-acyltransferase (cute)"
/protein_id="AAD07250.1"
/db_xref="GI:2313269"
/transl_table=11
/translation="MRLILFNQNAFLACMEVSSVYVNAVLDAVAIENPYISITLSTL
LAPLSMLAPLKPTRNSAFALGFVQALILFYCALSFYSDFYLLPLIIVLALVGV
LPLVLYENPYERLSPFGSSFIHPFGDMVLPDSFYSVFRVDKSLGLVFLAGV
LSTPKLKKVRIIGVILLGALDPNGPSTLKKVGNIELVSTKTPOLKPPDSVIND
IENNLIKETIKAGQKQTLVYPPERAYIALENSFPKAKLELDSNITAILIGTATOG
YNLNSFLFSKESVQINDKVIPLAPFGSTPLPEFLQKLELFGESYILRNAPHF
SDPTLIDTTPPLLCYESTGSPASNSPSKFIYVSNNAWSPSIEPTLQRTLLKYA
RKYDITLHSANFSTSYLSPSLIGLILFRRS"
6378.7052
/gene="HP0181"
6378.7052
/gene="HP0181"
/feature="hypothetical protein; identified by Genemark;"

```


source

1. .2006

/organism="Campylobacter jejuni"

/mol_type="genomic DNA"

/strain="81-176"

/db_xref="taxon:197"

94. .915

/gene="cbf2"

94. .915

/gene="cbf2"

/codon_start=1

/transl_table=11

/product="cell binding factor 2"

/protein_id="CA59175.1"

/db_xref="GI:671840"

/db_xref="GOA:O46105"

/db_xref="UniProt/Swiss-Prot:Q46105"

/translation="MKKFSVAATLIACVILNVNATVATVANGKISDTEVESEFPAPLRGODFKLPNOKKALIQOYIMODLIIDAKKOKLEKOPLYTEKELDAKDALVNVNOKELINTIKIDAAKFAFYDONKDKVYAPVQAKHIIIVATEKEAKDIIINELKGLKGGELDKFSELAKEKSIDPSSKNOGGEIEMFDOSTWKPPTDAFALKNTITTPKXTTFCYHIVILEKNSQAKQIKFDEVKOSITENGLKFEFKKVIYNOGQDLNSAKVEYK"

915. .1979

/gene="fdac"

915. .1979

/gene="fdac"

/codon_start=1

/transl_table=11

/product="fructose 1,6-bisphosphate aldolase"

/protein_id="CA59176.1"

/db_xref="GI:671841"

/db_xref="GOA:P53818"

/db_xref="UniProt/Swiss-Prot:P53818"

/translation="MGVLDIVAGVISDDELNKTIDYAKAEQFAIPAVNVGTDSINFVLEAKNSAEVPIIQPSNGKAFYAGKRCNGEIVGAIISGAKVHLLAKAGVPIILHIDHAKRLLPIMIDGLEIANAOYKFKTHGOALPSSHMLDSESLSEENISTCEVYLQKDLALGVALLEIPEICTGGEEDGVNNTGIDNSKLTQPEDVALAYERLGKISDKSIIASFQNVHGVGVNYSLOPEILKNOSOKFYKDKFALNSDKPINFVPHGSGSLIKQIKAVSGVTKANRDTPIQKAFWDCVREYELKNRAYILOGQIQINPEGDDKPNKCTYDPRVWLRSGGESMKIRLEIAEDINLCINKN"

gene

CDS

Db	Accession	Gene	Protein	Length (aa)	Accession	Gene	Protein	Length (aa)
Db	AACTAATGAAAAAATTTTCTTTAGTCGCAGCACTTGGATGAGGCTGTAATTAAAT	147	147	147	Db	AACTAATGAAAAAATTTTCTTTAGTCGCAGCACTTGGATGAGGCTGTAATTAAAT	147	147
Qy	27	AlaAsnAlaThrHisAsnThrLysLeuThrThraSerSerAlaGluValLeuAla	46	46	Qy	27	AlaAsnAlaThrHisAsnThrLysLeuThrThraSerSerAlaGluValLeuAla	46
Db	148	GTAATATGAGCTACA-----GTAGCT	168	168	Db	148	GTAATATGAGCTACA-----GTAGCT	168
Qy	47	ThrValAspGluArgProIleThrLysSerAspPheAsp-----MetIle	61	61	Qy	47	ThrValAspGluArgProIleThrLysSerAspPheAsp-----MetIle	61
Db	169	ACTGTTAATGCGAAGACATATGCGAATACAGTAAGTAAGTAATTTTGGCCCTATGCTT	228	228	Db	169	ACTGTTAATGCGAAGACATATGCGAATACAGTAAGTAAGTAATTTTGGCCCTATGCTT	228
Qy	62	LysGluIlnArgAsnProAsnPheAspPheAspLysLeuLysGluLysGluValLeu	81	81	Qy	62	LysGluIlnArgAsnProAsnPheAspPheAspLysLeuLysGluLysGluValLeu	81
Db	229	AGAGACAGCAG-----GATTTTAAACCTTGGCAGATTAATCAAAAAAGCTCTT	276	276	Db	229	AGAGACAGCAG-----GATTTTAAACCTTGGCAGATTAATCAAAAAAGCTCTT	276
Qy	82	IleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGluLysLeuAsp	101	101	Qy	82	IleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGluLysLeuAsp	101
Db	277	ATTGACGCAATATATATATGCAAGATTTTAATTTGCAATGCTATAAAAAACAATTTTGA	336	336	Db	277	ATTGACGCAATATATATATGCAAGATTTTAATTTGCAATGCTATAAAAAACAATTTTGA	336
Qy	102	SerThrProGluPheLysValMetMetGluAlaValLysLysGlnAlaLeuValGluPhe	121	121	Qy	102	SerThrProGluPheLysValMetMetGluAlaValLysLysGlnAlaLeuValGluPhe	121
Db	337	AAAGACCTTTATGACAAAAAGAACTGATCGTCGCAAAAAGATGCAATATCTGTTAAAGTT	396	396	Db	337	AAAGACCTTTATGACAAAAAGAACTGATCGTCGCAAAAAGATGCAATATCTGTTAAAGTT	396
Qy	122	TryAlaLysLysGlnAlaGluGluValLysLysValGlnIleProGluLysGluMetGln	141	141	Qy	122	TryAlaLysLysGlnAlaGluGluValLysLysValGlnIleProGluLysGluMetGln	141
Db	397	TATCAAGAGAAATTTTAAATTAATCTATTAA-----ATTGATGCGCTAAAGTTAA	447	447	Db	397	TATCAAGAGAAATTTTAAATTAATCTATTAA-----ATTGATGCGCTAAAGTTAA	447
Qy	142	AspPheGyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHis	161	161	Qy	142	AspPheGyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHis	161
Db	448	GCTTTTATATGATCAAAATTAAGACAAATATGTAACCACTGCAGAGTCGCAAGCAAAACAT	507	507	Db	448	GCTTTTATATGATCAAAATTAAGACAAATATGTAACCACTGCAGAGTCGCAAGCAAAACAT	507
Qy	162	IleLeuValLysThrGluAspGluAlaLysArgIleIleSerGluIleAspLysGlnPro	181	181	Qy	162	IleLeuValLysThrGluAspGluAlaLysArgIleIleSerGluIleAspLysGlnPro	181
Db	508	ATCTTAGTAGCGACAGAAAAAGAAAGCTAAGAGATATTAATTAACGAATTAAGGTTTAA	567	567	Db	508	ATCTTAGTAGCGACAGAAAAAGAAAGCTAAGAGATATTAATTAACGAATTAAGGTTTAA	567
Qy	182	LysAlaLysLysGluLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsn	201	201	Qy	182	LysAlaLysLysGluLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsn	201
Db	568	GGTAAACAACTAGATGCTAAATTTAAGGACCTGCTAAAGAAATCAATTTGATCCAGCT	627	627	Db	568	GGTAAACAACTAGATGCTAAATTTAAGGACCTGCTAAAGAAATCAATTTGATCCAGCT	627
Qy	202	SerLysAsnAlaGlnAsnGlyLysAspLeuGlyLysPheGlnLysAsnGlnMetAlaPro	221	221	Qy	202	SerLysAsnAlaGlnAsnGlyLysAspLeuGlyLysPheGlnLysAsnGlnMetAlaPro	221
Db	628	TCAAAAAAC-----CAAGGTGGAGACCTGGTTGGTTGATCAATCAACTATAGGTAAG	681	681	Db	628	TCAAAAAAC-----CAAGGTGGAGACCTGGTTGGTTGATCAATCAACTATAGGTAAG	681
Qy	222	AspPheSerLysAlaAlaPheAlaLeuThrProGluAspPyrThrLysThrProValLys	241	241	Qy	222	AspPheSerLysAlaAlaPheAlaLeuThrProGluAspPyrThrLysThrProValLys	241
Db	682	CCTTTTACAGATGCTGCTTTCGGCCTTAAAAAGTACTATTAACAACCTCGCGTAA	741	741	Db	682	CCTTTTACAGATGCTGCTTTCGGCCTTAAAAAGTACTATTAACAACCTCGCGTAA	741
Qy	242	ThrGluPheGlyIleThrIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle	261	261	Qy	242	ThrGluPheGlyIleThrIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle	261
Db	742	ACGAAATTTGGTATCATGTAATCTTAAAGAAATTCGCAAGCTAAAGCTCAATCA	801	801	Db	742	ACGAAATTTGGTATCATGTAATCTTAAAGAAATTCGCAAGCTAAAGCTCAATCA	801
Qy	262	TyrGluGlnAlaLysProThrIleLysGlyMetLeuGlnGluLysLeuPheGlnGluArg	281	281	Qy	262	TyrGluGlnAlaLysProThrIleLysGlyMetLeuGlnGluLysLeuPheGlnGluArg	281
Db	802	TTTATATAGTAAGTAACAAGATATTTGAACCGAATTTAAATTTGAAGATTTTAAAAAGTT	861	861	Db	802	TTTATATAGTAAGTAACAAGATATTTGAACCGAATTTAAATTTGAAGATTTTAAAAAGTT	861
Qy	282	MetAsnGlnArgIleGluGluLeuArgLysHisIleLysIle	295	295	Qy	282	MetAsnGlnArgIleGluGluLeuArgLysHisIleLysIle	295
Db	862	ATCATATCAAAAGCCAAAGATCTTTAAATAGTCTTAAAGTG	903	903	Db	862	ATCATATCAAAAGCCAAAGATCTTTAA	

TITLE Basham,D., Chillingworth,T., Davies,R.M., Feltham,T., Holtrope,S., Jagsis,K., Karlyshev,A., Moule,S., Pallen,M.J., Penn,C.W., Quail,M., Rajadream,M.A., Rutherford,K.M., Vanlilet,A., Whitehead,S. and Barrell,B.G.
The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences

JOURNAL Nature 403 (6770), 665-668 (2000)

MEDLINE 20150912

PUBMED 10688204

REFERENCE 2 (bases 1 to 308601)

AUTHORS Parkhill,J.

TITLE Direct Submission

JOURNAL Submitted (09-FEB-2000) Submitted on behalf of the Campylobacter sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

COMMENT Notes:
Details of *C. jejuni* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/C_jejuni/).

FEATURES
source
1. 308601
/organism="Campylobacter jejuni subsp. jejuni NCTC 11168"
/mol_type="genomic DNA"
/strain="NCTC 11168"
/db_xref="taxon:192222"
/complement(101..886)
/gene="Cj0343c"
complement(101..886)
/gene="Cj0343c"
complement(101..886)
/note="Cj0343c, probable integral membrane protein, len: 261 aa; similar to many hypothetical membrane proteins e.g. Y441 METUA M00441 (267 aa), fasta scores; opt: 152 z-score: 301.8 E(): 1.5e-09, 30.7% identity in 257 aa overlap. 45.1% identity to HP0677 and 44.0% identity to HP0226"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB74180.1"
/db_xref="GI:6967818"
/db_xref="UniProt/TREMBL:Q9PIF7"
/translation="MTIDLPYLIIIGISIASGLFGIGGNIIVPSMFALGASAHNA IGLSVLQMIFAVFGSYINKKQNLNDGIMIGGLIGASFGMLKALSDVALTS VGLASCIFFIKYAFGIKENIYVONRSWVKVILFAGAFGIPALISIGGGLLIA PLAYELIGDSKKVVSLSFVIFASVSGIISFNSGVIDSEVIHKGLVIGASVWGV FIGIKIRKMHSAHKRKLICVYALSIIGTHSLNKNLFINP"
1009..1134
/gene="Cj0344"
1009..1134
/gene="Cj0344"
/note="Cj0344, unknown, len: 41 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj0344"
/protein_id="CAB74181.1"
/db_xref="GI:6967819"
/db_xref="UniProt/TREMBL:Q9PIF6"
/translation="MFQNIKKYKDRILIFLNKQVLYLLIKINIDFKNFKHSLN"
1299..2549
/gene="trpE"
1299..2549
/gene="trpE"
/EC_number="4.1.3.27"
/note="Cj0345, trpE, possible anthranilate synthase component I, len: 416 aa; similar in C-terminal half to many e.g. TRPE_C10TM anthranilate synthase component I (EC 4.1.3.27) (494 aa), fasta scores; opt: 784 z-score: 1209.9 E(): 0.47, 0% identity in 281 aa overlap. 35.6% identity in 365 aa overlap to HP1282. Contains Pfam match to entry PF00425 chorismate bind, chorismate binding enzyme, score 309.50, E-value 3.9e-89"
/codon_start=1
/transl_table=11

/product="putative anthranilate synthase component I"
/protein_id="CAB74182.1"
/db_xref="GI:6967820"
/db_xref="GOA:Q9PIF5"
/db_xref="UniProt/TREMBL:Q9PIF5"
/translation="MFVKNVNFYRQILIEKENSYPADLTIRKVIIGIDCDYDANELS FSEFKAQYKALSKNKICDPAGFPGVPSANVSLFEKIPLSKKNYDPLPLFANAKA YLIEKNSKMFEGKASCYFEYLKDDIEPMKKQKNDPEIINSLEDDNDFLMEKKA KEYLISGDIPOVVISKOLCIGKOVNAFVYSLSALNSPAAVMYFSPKYGVLGSSPE FLIKIKREIYIAPLAGTRNLNNNDILALDKDLSDSEKLSSEHKMLVDLARNDSKPG GTQTRVENTESITIKKPMYMHVSEYRANMKDASIFDYIEAVFPAGTISGAPKIRALE ITSLEHEDCKRGIYGAGVFLENEDITLILIRCAFPYQDAVILASGAGIVLQSSQK EYAEICAKRKLALVAFENLKKNQ"
1743..2522
/gene="trpD"
/note="Pfam match to entry PF00425 chorismate bind, chorismate binding enzyme, score 309.50, E-value 3.9e-89"
2546..4147
/gene="trpD"
2546..4147
/gene="trpD"
/EC_number="4.1.3.27"
/note="Cj0346, trpD (trpGD), probable anthranilate synthase component II, len: 533 aa; (contains glutamine amidotransferase and anthranilate phosphoribosyltransferase) similar to many e.g. TRPG_BCOLI anthranilate synthase component II (EC 4.1.3.27) (530 aa), fasta scores: opt: 866 z-score: 781.3 E(): 0.32, 6% identity in 528 aa overlap. N-terminus has 34.6% identity to HP1281, C-terminus 16.33, 2% identity to HP1280. Contains PF00442 Glutamine amidotransferases class-I active site and Pfam matches to entry PF00117 GATase, glutamine amidotransferases class-I, score 252.00, E-value 8.4e-72 and to entry PF00591 Glycosyl transferase family, score 138.60, E-value 1.1e-37"
/codon_start=1
/transl_table=11
/product="anthranilate synthase component II"
/protein_id="CAB74183.1"
/db_xref="GI:6967821"
/db_xref="GOA:Q9PIF4"
/db_xref="UniProt/TREMBL:Q9PIF4"
/translation="MTILLIDYDFVFNKSMELBOLSNDEILVRNDALISLSEIKNLN PTHILISBPGHPGSGICLEIFPARINIPVLGICLQIOLALAFDSLVVQKQPMHA KNSILIKOCRENELESNLPSNFSVRKHSLEKQSLDEILALDEKGYMLAQNDPQ VYGVQRPDSYFSEYVGLQLENSPLKQDIKKQKQENPLSYLQKMSSEHIDIGFQ ICKTIRSKDYEILQVVALILITKSKLSKLSLAFVQILRYSGTQPSDESMIDICGT GGDFPKSINVSATSAFLIAGVAVAKGNNAISSSGSTDLALNITTPNLTESVL KQLNNQGLSFLHAPFPHVLGELKELRSRLGVRFVNLGFLHPNLKLYQLMGNYH APVHRLILEVKNIGRKALKVVRNGNDGIDELSDSEIKYELCGEILLEYISICEQFG FKRAHSEIIGSSAYDNAKDKDILSGRMQAKRDLVLTANMFALYTANKKSSPLVAK DMILEAIYSGNVIYFKEYQAYAA"
2546..3103
/gene="trpD"
/note="Pfam match to entry PF00117 GATase, glutamine amidotransferases class-I, score 252.00, E-value 8.4e-72"
2765..2800
/gene="trpD"
/note="PS00442 Glutamine amidotransferases class-I active site"
3143..3589
/gene="trpD"
/note="Pfam match to entry PF00591 Glycosyl transferase family, score 138.60, E-value 1.1e-37"
4134..4733
/gene="trpF"
4134..4733
/gene="trpF"
/EC_number="5.3.1.24"
/note="Cj0347, trpF, probable N-(5'-phosphoribosyl) anthranilate isomerase, len: 199 aa; similar to many trpF e.g. TRPF_LACCA

Qy	321	GIUGIGIYVRGHIshpRcophenethzsnitYVRGPGGInpethYValaRgThr	340
Db	106823	GAAAGCGACGACACTCCATTCTTCAATGATACCGACTCAGTTCTATGTTAGACT	106882

Best Local Similarity:	94.22%	Conservative:	17
Best Global Similarity:	89.95%	Mismatches:	22
Query Match:	89.65%	Indels:	1
		Gaps:	1

DB:

```

/organism="Molliella succinogenes"
/mol_type="genomic DNA"
/strain="DSM 1740"
/db_xref="taxon:844"
1..1203
/gene="tuf"
/gene="tuf"
1..1203
/gene="tuf"
/codon_start=1
/transl_table=11
/product="elongation factor Tu"
/protein_id="CA55419.1"
/db_xref="GI:587590"
/db_xref="GON:P42482"
/db_xref="UniProt/Swiss-Prot:P42482"
/translation="MAKKKFKKKRPVNIIGTIGHVDHGKTTLSAASIAVLTAKLCET
KDYADIDNAPBERERGITATSHIEVETERNRYAHVDCPEHADYVKNMIGCAQMDGSG
LIVSAADPMPEQTRHSILTSIQGVPIYVFLNKEDMVDADIELLEVMESVRLSL
YDEPGDPTIVGSAKALAEANDQNVGSGEKVYKLAENVRYIPTPEADVDKPF
MEVEFVETISAGTGVATGIVTEGVVAVGVGEVEIVGIIRNOKTIVTGVEMFPEKLEKDE
AGDNYVILRGKREKPEVSGMVLCTGISTPHNPEGEVYVLSKEGGRHTRPPNGVGR
PQPTVITTDVGTISISLPEGVENVMPGDNKNINVELLAIPALBEGSTRPIBEGGTVGA
GVVETIKR"

```

US-10-039-183A-4 (1-399) x WS1740TUF (1-1203)

QY 1 MetAlaLysGluYbPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
 Db 1 ATGGCTGAAGAAAAGTTGTGTAATCAAAACCCACGTTAAATATGCGTACATCGGTGC 60
 QY 21 ValAspHisGlyLysThrThrLeuSerAlaIleSerAlaValLeuSerLeuYsgly 40
 Db 61 GTTGAACACGCTAAACCACTCTAGTCCGCTATTTTGGCGGTACTTGCACCAAGT 120
 QY 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArggly 60
 Db 121 CTTTCGAGCTTAAAGATTATGATGCGATCGACAATGCTCTCGAAGAGAGAGAGCGTGT 180
 QY 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisGlyTyrAlaHisVal 80
 Db 181 ATCACCACATCGCTACTCACACATCGAGTATGAACAGAAATCGACACTACGCTACGCT 240
 QY 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaIleGluMetAsp 100
 Db 241 GACTCCCTCGACACGCGCATCTATGTAAACATGATTAACAGTCTCTCAATGAT 300
 QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
 Db 301 GCGCGGATCTGTGTGTTCTCGCGGAGTGGCCCATGCCCAACTAGGAGACAT 360
 QY 121 LeuLeuSerArgGluValGlyValProHisIleValIlePheLeuAsnLysGluAspMet 140
 Db 361 CTTCTTTCTCGACAAGTAGGCGCTCTTACATCGTGTGTTTCTTGAACCAAAAGATAG 420
 QY 141 ValAspAspGluGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
 Db 421 GTTGAATGACGCTGAGCTTCTTGAAGCTTGTAAATGGAAGTATGAGAACTTCTAGCAAC 480
 QY 161 TyrGluPheProGlyLysAspSerThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
 Db 481 TACGACTTCCCTGAGATGACACTCTCATCGTTGAGGTTCCGCTTAAAGCTCTTGAA 540
 QY 181 GluAlaLys--AlaGlyAsnValGlyGluTyrProGlyLysValLeuLysLeuMetAla 199
 Db 541 GAGGCTAAACGACCAAGAAATGTGGCGAGTGGCGGAGAAAGTATGAAAGCTTATGGT 600
 QY 200 GluValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetPro 219
 Db 601 GAGGTGACCGATATATCTTACGCTGACGAGATGCGAATGCGAATGCGTCTTATGCT 660
 QY 220 ValGluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArg 239
 Db 661 GTTGAAGACGTATCTCCATCGCGGTCGTGGAACGTTGTGACAGAAAGATTGAAAGA 720
 QY 240 GlyValValLysValGlyAspGluValGluIleValGlyIleArgProThrGluLysThr 259
 Db 721 GCGGTGTAAAGTGGTGAACGAGTGAAGAACTGTTGATTCGAAACACACAAACAA 780
 QY 260 ThrValThrGlyValGluMetPheArgLysGluLeuGluLysGlyValAlaGlyAspAsn 279
 Db 781 ACCGTAATCGCGGTGAGATGTTCCGAAAGAGCTCGACAAAGGCGGTGACAAAC 840
 QY 280 ValGlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuGly 299
 Db 841 GTTGGTGTCTTTTGAAGGACCAAGAAAGAAAGTGTGAGAGAGATATGCTTCTTGT 900
 QY 300 LysProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLys 319
 Db 901 AAAATATGTTCTATCACTCTCACATTAATTGAAGGAGGATTACCTTCTTCCAA 960
 QY 320 GluGluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGluPheTyrValArg 339
 Db 961 GAGGAGGCGGACGACACATCTTCAATGATGATACCGACTCAGTTCTATGTTAGA 1020
 QY 340 ThrThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGly 359

Db 1021 ACTACAGCGTTACCGGTTCTATCTCTCTCGAGGCGGTAGAGATGTTATGCTGT 1080
 QY 360 AspaenValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrLys 379
 Db 1081 GACACGTTAAAGATTAATGTTAGCTTATCGCTCTGAGCCCTGAAAGAGGAAACAGA 1140
 QY 380 PheAlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIle 397
 Db 1141 TTCCGATCCCGTGAAGGTGTGTGAACCGTTGGTGGGAGTGTGTTACCAAGATC 1194
 RESULT 8
 CUY17167
 LOCUS
 DEFINITION
 genes.
 ACCESSION
 Y17167.1 GI:3114760
 KEYWORDS
 EF-Tu; transfer RNA-Gly; transfer RNA-Thr; transfer RNA-Tyr;
 tRNA-Gly; tRNA-Thr gene; tRNA-Tyr gene; tufB gene.
 SOURCE
 ORGANISM
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 Campylobacteraceae; Campylobacter.
 REFERENCE
 1 Thies, F. and Giegerich, G.
 JOURNAL
 Unpublished
 2 (bases 1 to 1865)
 REFERENCE
 Giegerich, G.
 TITLE
 Direct Submission
 Submitted (22-APR-1998) G. Giegerich, Department of Neurology,
 University of Regensburg, Universitaetsstr. 84, D-93053 Regensburg,
 FRG
 FEATURES
 source
 1. location/Qualifiers
 1. 1865
 /organism="Campylobacter jejuni"
 /mol_type="genomic DNA"
 /db_xref="taxon:197"
 100. 176
 /gene="tRNA-Thr"
 100. 176
 /gene="tRNA-Thr"
 /product="tRNA-Thr"
 225. 311
 /gene="tRNA-Tyr"
 225. 311
 /product="tRNA-Tyr"
 317. 394
 /gene="tRNA-Gly"
 317. 394
 /product="tRNA-Gly"
 498. 573
 /gene="tRNA-Thr"
 498. 573
 /product="tRNA-Thr"
 632. 1840
 /gene="tufB"
 632. 1840
 /gene="tufB"
 641. 1840
 /gene="tufB"
 /codon_start=1
 /transl_table=11
 /product="EF-Tu protein"
 /protein_id="CAA7676.1"
 /db_xref="GI:3114761"
 /db_xref="GOA:O69303"
 /db_xref="UniProt/Swiss-Prot:O69303"
 /translation="MAKERSRNKPHVNIIGTIGVDDHCKTTLTAALSAVLSRRGLAEI
 KDYNIDNAPERKERGITIATSHLYEIEDNDNVAHVDCPGADYVKKNTTGAQMDGA
 ILVVSADGPMFQTRHILLRSQVGPYIVFVNKKADVDAELLELEVMERILLSS

Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Web site. (URL: <http://www.astraboston.com/hpy1ori>).

FEATURES

source
1.10085
/organism="Helicobacter pylori J99"
/mol_type="genomic DNA"
/strain="J99"
/db_xref="taxon:85963"
complement(235..708)
/gene="jhp0153"
complement(235..708)
/note="similar to H. pylori 26695 gene HP0167"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05736.1"
/db_xref="GI:4154669"
/translation="MAGRMELKNITSETLNEIKMAKTIDGENTAKTSPFPTPSH
LONTPNSKANPTEPKNAKETEOTEKEEKEKEITTEIVQENPNQALIPERV
FLKGLBERTLVFKGMQALEKEALRDLVAFIQYQLSVLEKLESLENERNT"
complement(687..950)
/gene="jhp0154"
complement(687..950)
/note="similar to H. pylori 26695 gene HP0168"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05737.1"
/db_xref="GI:4154670"
/translation="MPLRTITLAHIYEEGPFPEALQITNTILKXTPDNTALKKMR
LEKIQNCAPFHDALIKOHYINFTKGRDLSVENLEKMLVEN"
complement(997..2265)
/gene="jhp0155"
complement(997..2265)
/note="similar to H. pylori 26695 gene HP0169"
/codon_start=1
/transl_table=11
/product="putative PROTEASE"
/protein_id="AAD05738.1"
/db_xref="GI:4154671"
/translation="MNQVELSPAGNIKKLTALNYGADAVYGVSHPSIRNACKEP
TLTRREGIDIAHALDKKYATATNGPENSQIKLBEHLYKAELBDPFTIAAGVI
KLASKIAPHIVLSTQAVNTLDAQVYDGVRIYCARLSLNDAAEIKKALPDL
ELIEFVHGSMCFPSGRCILSALOKGRVNRSCANDCFEYEVYKQNDVMMRLV
EEEGVTHIFNAKDNLGSHLAEISSNALSALKTEGRTKSYAAQTIRLRLVLD
FYHNTLKPSFVSELTNKGRTGYLMRPFEDLDQNHOTALSEGFOVNGEITTE
DGRFACKFTTTNTAYELIAKMAITPIVNDIKITTEFGKSLVLYKILLENTTE
LETISGNVNLRLPAPLPAPSFLLRTQVRV"
complement(2268..2999)
/gene="jhp0156"
complement(2268..2999)
/note="similar to H. pylori 26695 gene HP0170"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05739.1"
/db_xref="GI:4154672"
/translation="MTQBELDALMSGGLENLBALAKEAKEPTKEEPKEKENSNT
SEKMTVKEDAKYKISPHENPPPEEHEVHQLDVTSDSEVKATQIPDQDLI
GASAEIKAMVKIKQIEPLQKQIEFDNLHAPPHVESFRTALNQEOLINAKSLIEE
AANCSDSSWQAMDIMQFODIHRKITERVVNVRALSQVNSLFEKIDSKSVSATY

gene
complement(3069..4160)
/gene="prfb"
/note="synonym: jhp0157"
complement(3069..4160)
/gene="prfb"
/note="similar to H. pylori 26695 gene HP0171"
/codon_start=1
/transl_table=11
/product="PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2)"
/protein_id="AAD05740.1"
/db_xref="GI:4154673"
/translation="MDNYTSELKSLONKCDNALIIPKEIKQELERIEKEQBDPN
FMQDVLKARDNKEKVRNRLLETQOKKNSLDSEVELFELAQNDSEVTLSLEYEA
PTLHSHVQKEIEMLSGENDASNALITIQGAGTESQDMASILRYMLNRAEKSF
KSEILDYODGEAGIKGVAFLIKGENGAYGLKNENGVRILVIRISPDANAKRHSFAS
VOISPELDDIDIDEKQVRYDYIRASAGAGQHNTKESAVRITHTPPTGVQCCND
RSQHNKASALMKSKULELELEKQSSAKNEKSEIGWGHQIRSYVLAPYQOVKDA
RSNIAYSNVEAILDGDIDAILDEGVLIARA"
complement(4221..5393)
/gene="moea"
/note="synonym: jhp0158"
complement(4221..5393)
/gene="moea"
/note="similar to H. pylori 26695 gene HP0172"
/transl_table=11
/product="MOLYBDOPTEIN BIOSYNTHESIS PROTEIN"
/protein_id="AAD05741.1"
/db_xref="GI:4154674"
/translation="MISFEKALKIHSNIPKLEVEVVSLESAGRILIAEDIIYHAL
PKFNQSAMADGYGFKQDQGTQVIOHI FADGDVSALVEKNECVKMTGMAVPGIE
TIVPECLSHKQDLAPKDFKIHANROGENASLNSVLPKMTRLNYGHILIAS
QGFKEIKARRLKIALFSSGDEIVLQGNALCEQVYDVSNGVFMMLKNYTHFGLV
KDDKRLQIKITIELOGDYITLSSAGVSDKQPFYDALKERALLFYEEKYNLKPGEPVT
LAQLQNSITIGPGNPISCLVLRVLIPLVLRSLNDQFKLKPKAQINAPKLNK
RTHILIGVSNHGFIPYNNRYESGAILQALQAVDSIALIDEGVLQGSIEILRFEN"
complement(5403..6170)
/gene="flir"
/note="synonym: jhp0159"
complement(5403..6170)
/gene="flir"
/note="similar to H. pylori 26695 gene HP0173"
/codon_start=1
/transl_table=11
/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
/protein_id="AAD05742.1"
/db_xref="GI:4154675"
/translation="MLDFIOELSTPHVRDFLLPLRVSGVLSFPFPFENHLVPLSVRG
ALSILYSAIFPYTLFESNAIYTBEGFIACCELFQVCASVFIQIVASIVFATDSI
SFSMGLTMASAVDPISGSOKPIVGGALLLAILLLDPSFHHQIILFPDHSIKAVPLG
RFPVEPELAKNIIVAFSHLPIKESMAPRILICIVLISPIIIGIMKTHPQNFLLAIGF
PVKAIAGVGIILIASAIMGRPEKISLAFYSIKIP"
complement(6164..6892)
/gene="jhp0160"
complement(6164..6892)
/note="similar to H. pylori 26695 gene HP0174"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05743.1"
/db_xref="GI:4154676"
/translation="MTVFLSPKFSFNDPLSABMILINLSPILSLAFPGAIFYNNG
STVNCQTLFQSLNDVAHSGGFRAGVFAWFKALVYFLIRWYIYLLSLVINIPASIF
YTPLVASYLHQKYYPHVVLBERGSLFSEIKYFLKSLTLMLLFLAVLTPLVLPPIGVF
GVFSPSIVPHPLFFKNTMSLIDIASMIFNNQSYONLKKORLKHRRSPFCYLSLIPF
NPFATLQTLMLTHTYFPIFKKEKC"
761..8260
/gene="jhp0161"
761..8260
/gene="jhp0161"
/note="similar to H. pylori 26695 gene HP0175"

Result No.	Score	Query Match	Length	DB	ID	Description
1	1520	100.0	1000	6	BD092341	BD092341 Ident:ff1
2	1520	100.0	1149	6	AR343398	AR343398 Sequence
3	1520	100.0	1149	6	ED082346	ED082346 76 kDa, 3
4	1520	100.0	11421	1	AE000538	AE000538 Helicobac

5	1486	97.8	10085	AE001454	AE001154 Helicobac
6	1482	97.5	1082	BD061716	BD061716 Helicobac
7	1482	97.5	1117	BD061705	BD061705 Antigenic
8	1474	97.0	1082	BD061965	BD061965 Antigenic
9	1311	86.2	844	AX788859	AX788859 Sequence
10	1017	66.9	616	BD061671	BD061671 Antigenic
11	664	43.7	595	AX790177	AX790177 Sequence
12	600.5	39.5	303820	AE011744	AE011744 Helicobac
13	482.5	31.7	349926	BX571660	BX571660 Helicobac
14	376.5	24.8	2006	CUGB22	CUGB22
15	376.5	24.8	308601	Cu11168x2	Cu11168x2
16	363	23.9	348971	BX572594	BX572594 Campylob
17	361.5	23.8	10179	AE009453	AE009453 Brucella
18	361.5	23.8	11118	AE014483	AE014483 Brucella
19	347.5	22.9	349498	AP003002	AP003002 Meserhizo
20	344	22.6	110000	AE011802_00	AE011802 Geobacter
21	333.5	21.9	340900	SME551180	AE1591791 Sinorhizo
22	329	21.6	300600	AP005935	AP005935 Bradyrhiz
23	328	21.6	100000	AE011802_26	AE011802 Clostridi
24	322	21.2	13315	AE007541	AE007541 Clostridi
25	320	21.2	110000	AE017282_17	AE017282 Clostridi
26	317	20.9	300029	AE015936	AE015936 Clostridi
27	315.5	20.8	302300	AP003186	AP003186 Clostridi
28	303	19.9	10765	AE008331	AE008331 Agrobacte
29	300	19.9	11689	AE009281	AE009281 Agrobacte
30	300	19.7	110000	BX8997700_02	BX8997700 Clostridi
31	295	19.4	35996	AY458645	AY458645 Unculture
32	294	19.3	110000	BX897695_02	BX897695 Clostridi
33	288.5	19.0	10707	AE013197	AE013197 Thermosua
34	278.5	18.3	305584	AE015920	AE015920 Chromobac
35	269.5	17.7	298050	BX321861	BX321861 Nitrosomo
36	264.5	17.4	110000	CP000010_14	CP000010 Clostridi
37	264	17.4	304708	AE017001	AE017001 Bacillus
38	261	17.2	288727	AE017268	AE017268 Bacillus
39	260.5	17.1	110000	CP000001_11	CP000001 Clostridi
40	259.5	17.1	110000	BX571965_16	BX571965 Clostridi
41	258	17.0	110000	AE001355_11	AE001355 Clostridi
42	257.5	16.9	12746	AE008642	AE008642 Rickettsi
43	256	16.8	299550	AP001511	AP001511 Bacillus
44	255.5	16.8	1242	RPR219071	RPR219071 Rickettsi
45	255.5	16.8	1644	RPDNAMEBP	RPDNAMEBP Rickettsi

CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source Location/Qualifiers
1..1000
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Alignment Scores:
Pred. No.: 1.13e-108 Length: 1000
Score: 1520.00 Matches: 299
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 6

US-10-039-183A-2 (1-299) x BD092341 (1-1000)

QY 1 MetLysLysAsnLLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 51 ATGAAAAAAATATCTTAATTTAGCGTTAGTGGGCGTTGAGCAGTCGTTTGTGATG 110
QY 21 AlaLysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrThrAspSer 40
DB 111 GCTAAGCCGCGCTCATTAAGCAATAACGCTACGATTAACGAAAAAAGCACTGATTC 170
QY 41 SerAlaGlyValLeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMet 60
DB 171 TCAGCAGCGCGGTGTGCGCAGGTGATGCGACGCTATCATTAAGCAATTTTGCATG 230
QY 61 IleLysGlnArgAspProAsnPheAspPheAspLysLeuLysGlnLysGlnLysGlnAla 80
DB 231 ATTAAGCAACGAATCTTAATTTTGTGATTTTGACAGCTTAAGCAAGAAAGAAAGAGCC 290
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGlnLysLeu 100
DB 291 TTGATTTGATCAAGCTATTCGCAACCGCTTTGTAGAAATGAAAGCTAAACCGAATG 350
QY 101 AspSerThrProGlnLysLysAlaMetMetGlnAlaValLysGlnAlaLeuValGlu 120
DB 351 GACACGACCTCCAGAAATTTTAAAGCATGATGAAAGCGGTTAAAAAACAAGCTTTAGTGA 410
QY 121 PheTPAlaLysLysGlnAlaGlnGluValLysLysValGlnIleProGlnLysGlnMet 140
DB 411 TTTTGGGCTTAAGAAACAGCTGTAAGAGTGAAGAAAGTCCAAATCCCAAGAAAGAAAG 470
QY 141 GlnAspPheTyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArg 160
DB 471 CAAGATTTTTCACACGCTAACAAAGATCAGCTTTTGTCAAGCAAGAAAGCCCATGCTAG 530
QY 161 HisIleLeuValLysThrGlnAspGlnAlaLysArgIleIleSerGlnLysLysGln 180
DB 531 CATATTTTATGTAAGAAACCAAGATGAGGCTAAACGATTAATTTCTGAGATTGACAAACAG 590
QY 181 ProLysAlaLysLysGlnAlaLysPheIleGlnLeuAlaAsnArgAspThrIleAspPro 200
DB 591 CCAAGAGCTTAAGAAAGCAAGCTTAATTCATTGAGCTTACCAATCGGAGTACGATTGATCT 650
QY 201 AsnSerLysAsnAlaGlnAsnGlnLysAspLeuGlnLysPheGlnLysAsnGlnMetAla 220
DB 651 AACAGCAAGAAACGCGCAAAATGCGGCTGATTTGGGGAATTCCAAAAGAAACCAATGCGCT 710
QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlnAspTyrThrLysThrProVal 240
DB 711 CCGGATTTTTCATAAGCCGCTTTCGCTTAACCTCCGCGGAGATTACCTAAACCCCTGTT 770
QY 241 LysThrGlnLysPheGlnTyrHisIleIleTyrLeuIleSerLysAspSerProValThrTyr 260
DB 771 AAAACAGAGTTGGTTATCATATTAATTAATTTGATTTTCAAGATAGCCCTGTAACCTAT 830
QY 261 ThrTyrGlnGlnAlaLysProThrIleLysGlnMetLeuGlnGlnLysLeuPheGlnGln 280

DB 831 ACTTATGAACAGGCTTAACCTTACATTAAGGAGATTTACAGAAAGCTTTTCCAAAGA 890
QY 281 ArgMetAsnGlnArgIleGlnGlnLeuArgLysHisAlaLysIleValIleAsnLys 299
DB 891 GCGATGAATCAACGCGATTGAGGAGCTTAAGAAAGCAGCTTAATTTTATCAACAG 947

RESULT 2
AR342398 1149 bp DNA linear PAT 17-AUG-2003
LOCUS
DEFINITION Sequence 23 from patent US 6576244.
ACCESSION AR342398
VERSION AR342398.1 GI:33737371
KEYWORDS
SOURCE
ORGANISM
Unkown.
REFERENCE
1 (bases 1 to 1149)
AUTHORS Weltzin, R.A. and Guy, B.
TITLE LT and CT in parenteral immunization methods against helicobacter infection
JOURNAL Patent: US 6576244-A 23 10-JUN-2003;
FEATURES
source Location/Qualifiers
1..1149
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1.31e-108 Length: 1149
Score: 1520.00 Matches: 299
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 6

US-10-039-183A-2 (1-299) x AR342398 (1-1149)

QY 1 MetLysLysAsnLLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 106 ATGAAAAAAATATCTTAATTTAGCGTTAGTGGGCGTTGAGCAGTCGTTTGTGATG 165
QY 21 AlaLysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrThrAspSer 40
DB 166 GCTAAGCCGCGCTCATTAAGCAATAACGCTACGATTAACGAAAAAAGCACTGATTC 225
QY 41 SerAlaGlyValLeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMet 60
DB 226 TCAGCAGCGGTGTGCGCAGGTGATGCGACGCTATCATTAAGCAATTTTGCATG 285
QY 61 IleLysGlnArgAspProAsnPheAspPheAspLysLeuLysGlnLysGlnAla 80
DB 286 ATTAAGCAACGAATCTTAATTTTGTGATTTTGCAAGCTTAAGAAAGAAAGAAAGAGCC 345
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGlnLysLeu 100
DB 346 TTGATTTGATCAAGCTATTCGACCGCTTGTAGAAATGAAAGCTAAACCGAAGAAATG 405
QY 101 AspSerThrProGlnLysLysAlaMetMetGlnAlaValLysLysGlnAlaLeuValGlu 120
DB 406 GACACGACCTCCAGAAATTTTAAAGCATGATGAAAGCGGTTAAAAAACAAGCTTTAGTGA 465
QY 121 PheTPAlaLysLysGlnAlaGlnGlnLysLysValGlnIleProGlnLysGlnMet 140
DB 466 TTTTGGGCTTAAGAAACAGGCTGAAGAGTGAAGAAAGTCCAAATCCCAAGAAAGAAAG 525
QY 141 GlnAspPheTyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArg 160
DB 526 CAAGATTTTTCACACGCTAACAAAGATCAGCTTTTGTCAAGCAAGAAAGCCATGCTAG 585
QY 161 HisIleLeuValLysThrGlnAspGlnAlaLysArgIleIleSerGlnLysLysGln 180
DB 586 CATATTTTATGTAAGAAACCGAAGATGAGGCTAAACGATTAATTTCTGAGATTGACAAACAG 645


```

/codon_start=1
/translate=1
/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
/protein_id="AAD05744.1"
/db_xref="GI:4154677"
/translation="MKKNILNLTALVGLSASFLMAKPPAHNANSHTKETTDSAGV
LATVGRPTIKSDPMIKORNFDPDLKEKEKELIEQALRTLVNEAKAEKLNQ
TPEFKAMEAVKALVEFMKAKOAEVYKIOIPEKMODFYNAKDLFYQEAHAR
HLYKTEDEAKRIIEIDKOPKAKAEKFIELANDDTIPNSKNAQNSGDLCKPKNO
MAPDPSKAFALTPEDTYKTPKTEFGHIITLISKDSPVITYEQAPTKMLQEK
LFQERNQRIEELRHAKIVINK"
/gene="Eba"
/feature="synonym: jhp0162"
8274..9197
/gene="Eba"
8274..9197
/gene="Eba"
/feature="similar to H. pylori 26695 gene HP0176"
/codon_start=1
/translation="PRUCTOSE-BISPHOSPHATE ALDOLASE"
/protein_id="AAD05734.1"
/db_xref="GI:4154667"
/translation="MLVKGNEILLKAKEGYGVGFNFVNEFMALFEAGENSPL
FIOASEGAIKYMGIDMAVGMVKIMCERYPHIPVALHLDHGTTFESCEKAVKAGFTSV

```

Alignment Scores:

```

Pred. No.: 5,94e-105 Length: 10085
Score: 1486.00 Matches: 290
Percent Similarity: 99.00% Conservative: 6
Best Local Similarity: 96.99% Mismatches: 3
Query Match: 97.76% Indels: 0
Gaps: 0

```

US-10-039-183a-2 (1-299) x AE001454 (1-10085)

```

QY 1 MetLysLysAsnLLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
Db 7361 ATGAAAGAAAATATCTTAATTAGCGTTAGGGCGCGCTTAGCGCGCTTTTGATG 7420
QY 21 AlaLysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrThrAspSer 40
Db 7421 GCTAAGCCGGCTCATATGCGAATTAAGCTTACGATTAACAGAAAGAAAGCACTGATGCT 7480
QY 41 SerAlaGlyValLeuAlaThrValAsnGlyValArgProIleThrLysSerAspPheAspMet 60
Db 7481 TCAGACGGCGTGTGCGACAGTGTGACCAACCATCAACCAAAAGCATTTGATATG 7540
QY 61 IleLysGlnArgAsnProAsnPheAspPheAspLysLeuLysGlnLysGlnLysGlnAla 80
Db 7541 ATTAAGCAACGAAATCTTAATTGATTTTGACAAAGCTTAAAGAAAGAAAGAAAGAGCC 7600
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGlnLysLeu 100
Db 7601 TTGATTGAGCAAGCTATCGCACCGCGCTTGTAGAAATGAGGCTTAAGCAAGAAAGCTT 7660
QY 101 AspSerThrProGlnPheLysValMetMetGlnAlaValLysLysGlnAlaLeuValGln 120
Db 7661 AATCAGACTCCGAAATTTAAAGCGATGATGAGAAAGCGGTTAAAGCAAGGCTTTAGTGAA 7720
QY 121 PheTProAlaLysLysGlnAlaGlnLysValLysValGlnIleProGlnLysGlnMet 140
Db 7721 TTTTGGCGCTTAAAGAAAGCGCTGAAAGAGTGAAGAAAGATCCCAAGAAAGAAAGATG 7780
QY 141 GlnAspPheTyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArg 160
Db 7841 CAGGATTTTTCACAGCCCAATTAAGATCGCTTTTGTCAACAGCAAGAGCCCATCTAGG 7840
QY 161 HisIleLeuValLysThrGlnAspGlnAlaLysArgIleIleSerGlnIleAspLysGln 180
Db 7841 CATATTTTAAAGTAAACCGAAGATGAGCAAGCAAGAGATTTATTCGAGATTGACAAACG 7900
QY 181 ProLysAlaLysLysGlnAlaLysPheIleGlnLeuAlaAsnArgAspThrIleAspPro 200

```

```

Db 7901 CCAAGGCTTAAAGAAAGCAACCAATTCATGATGATGACCAATCGGATATGATGATCT 7960
QY 1 AsnSerLysAsnAlaGlnAsnGlnLysLysPheLysLysLysLysPheGlnLysAsnGlnMetAla 220
Db 7961 AACACCAAGAAAGCGCAAAATGCGGTATTTGGGAAATTCACAAAGCAAAATGCGT 8020
QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlnLysAspThrLysThrProAla 240
Db 8021 CCGATTTTTCACAAAGCGCGCTTCCCTTAACCTCGGGGATTAACATAAACCCCTGTT 8080
QY 241 LysThrGlnPheGlyTyrHisIleIleTyrLeuIleSerLysAspSerProValThrTyr 260
Db 8081 AAAACAGAGTTGGTTATCATATATATATGATTTGATTAAGATAGCCCTTAATCAT 8140
QY 261 ThrTyrGlnAlaLysProThrLysLysGlnMetLeuGlnLysLysPheGlnGln 280
Db 8141 ACTTATGAGCAAGCTTAAAGCTTACCTTAAGGGGATGTTACAAAGAAAGCTTTCCAGAA 8200
QY 281 ArgMetAsnGlnArgIleGlnLysLysLysLysLysLysLysLysLysLysLysLys 299
Db 8201 CGCATGATCAACGATGAGGATTAAGAAAGCAAGCTTAAATTTGATATCAACAG 8257

RESULT 6
BD061716 1082 bp DNA linear PAT 27-AUG-2002
LOCUS BD061716
DEFINITION Antigenic composition and method of detection for Helicobacter
Pylori.
ACCESSION BD061716
VERSION BD061716.1 GI:22607321
KEYWORDS JP 2001517091-A/50.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1082)
AUTHORS Chow,T.P., Fry,K.E., Lim,M.Y. and Mcatee,C.P.
TITLE Antigenic composition and method of detection for Helicobacter
JOURNAL Patent: JP 2001517091-A 50 02-OCT-2001;
GENELABS TECHNOLOGIES INC
COMMENT PN JP 2001517091-A/50
PD 02-OCT-2001
PF 25-APR-1998 JP 1998547263
PR 25-APR-1997 US 60/045107, 14-OCT-1997 US 60/061958 PI
THERESA P CHOW, KIRK E FRY, MOON Y LIM, C P MCATEE PC
C12N15/31, C07K14/205, C07K16/12, A61K39/106
CC Strandedness: Single;
CC Topology: linear;
FH Key Location/Qualifiers.
FEATURES
source 1..1082
location/Qualifiers
1..1082
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Alignment Scores:
Pred. No.: 1,06e-105 Length: 1082
Score: 1482.00 Matches: 291
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.32% Mismatches: 5
Query Match: 97.50% Indels: 0
Gaps: 0
US-10-039-183a-2 (1-299) x BD061716 (1-1082)
QY 1 MetLysLysAsnLLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
Db 165 ATGAAAGAAAATATCTTAATTAGCGTTAGGGCGCGCTTACGATGAGCGCGCTTTTGATG 224
QY 21 AlaLysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrThrAspSer 40
Db 225 GCTAAGCCGGCTCATATGCGAATTAAGCTTACGATTAACAGAAAGAAAGCACTGATTC 284
QY 41 SerAlaGlyValLeuAlaThrValAsnGlyValArgProIleThrLysSerAspPheAspMet 60

```

Db 285 TCACCCGGGCTGTAGCGACAGTGCAGACCTATCATTAAGCATTTTGATG 344
Qy 61 ILeYsGlnArGAsnProAsnPhaSpnPhaSpnLeuYsGlnuYsGlnuA 80
Db 345 ATTAAGCAACGAAATCCATATTGATTTGACAGCTTAAGAGAAAGAAAGAGCC 404
Qy 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnuAlaYerThrGluYsLeu 100
Db 405 TTGATTGAGCAACGCTATCCGACCGCAGCTGTAGAAAGAGAGCTTAAGCAAGAAAGCTC 464
Qy 101 AppSerThrProGluPhelYsAlaMetMetGluAlaValYsGlnAlaLeuValGlu 120
Db 465 GATCAGACTCCAGAAATTTAAACGATGATGAGAAAGCCGTTAAAGAAACAGGCTTAAGTGA 524
Qy 121 PheTPAlaYsLeYsGlnAlaGluGluValYsYsValGlnIleProGluYsGluMet 140
Db 525 TTTTGGGCTAAAGAAAGAGCTGAGAGAGTGAAGAAAGTCCAAATCCAGAAAGAAAGATG 584
Qy 141 GlnAspPheTYrAsnAlaAsnLYsAspGlnLeuPheValYsGlnuAlaHisAlaArg 160
Db 585 CAAGATTTTACACGCTTAATAAGATCAGCTTTTGTCAAGCAAGAACCCATGCTAG 644
Qy 161 HisIleLeuValYsThrGluAspGluAlaYsArgIleIleSerGluIleAspYsGln 180
Db 645 CATATTTTAGTAAAGAACCGAAGATGAGCTAAACGATATTCTGAGATTGACAAACAG 704
Qy 181 ProLYsAlaYsLeYsGlnuAlaYsPheIleGluLeuAlaAsnArgAspThrIleAspPro 200
Db 705 CCAAGAGCTTAAAGAAAGAACCAATTCATGAGTTAGCCAAATCGGATTCGATGATCTCT 764
Qy 201 AsnSerLYsAsnAlaGlnAsnGlyYsAspLeuGlyYsPheGlnLYsAsnGlnMetAla 220
Db 765 AACACAGAAAGACCGCAAAATGCGCGTGATTTGGGAAATTCCAAAGAAACCAATGGCT 824
Qy 221 ProAspPheSerLYsAlaAlaPheAlaLeuThrProGlyAspTYrThrYsThrProVal 240
Db 825 CCGGATTTTCTAAAGCCGCTTCCGCTTACCTTCGGGAGATTACACTAAACCCCTGTT 884
Qy 241 LysThrGluPheGlyTYrHisIleIleTYrLeuIleSerLYsAspSerProValThrTYr 260
Db 885 AAAACAGAGTTGGTTATCATATTATCTATTGATTCTTAAGATAGCCCTGTAACTTAT 944
Qy 261 ThrTYrGluGlnAlaYsProThrIleYsGlyMetLeuGlnGluYsLeuPheGlnGlu 280
Db 945 ACTTATGAGCAACGCTTAACCTACATTAAGGGATGTTTACAGAAAGCTTTTCCAGAA 1004
Qy 281 ArgMetAsnGlnArgIleGluGluLeuArgLYsHisAlaYsIleValIleAsnLYs 299
Db 1005 CCGATGAATCAACGCAATGAGAGATTAAAGAAAGCAAGCTAAATGTCTCAACAAG 1061

RESULT 7
BD061705 1117 bp DNA linear PART 27-AUG-2002
LOCUS Antigenic composition and method of detection for Helicobacter
DEFINITION Py1ori.
ACCESSION BD061705.1 GI:22607310
VERSION BD061705.1
KEYWORDS JP 2001517091-A/39.
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
1 (bases 1 to 1117)
REFERENCE
AUTHORS Chow, T. P., Fry, K. E., Lim, M. Y. and Mcatee, C. P.
TITLE Antigenic composition and method of detection for Helicobacter
JOURNAL Patent: JP 2001517091-A 39 02-OCT-2001;
GENELABS TECHNOLOGIES INC
PN JP 2001517091-A/39
COMMENT
PD 02-OCT-2001
PR 25-APR-1998 JP 1998547263
PR 25-APR-1997 US 60/045107,14-OCT-1997 US 60/061958 PT
THERESA P CHOW, KIRK E FRY, MOON Y LIM, C P MCATEE PC
C12M15/31, C07K14/205, C07K16/12, A61K39/106

CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source Location/Qualifiers
1..1117
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Alignment Scores:
Pred. No.: 1,1e-105 Length: 1117
Score: 1482.00 Matches: 291
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.32% Mismatches: 5
Query Match: 97.50% Indels: 0
Gaps: 0
US-10-039-183a-2 (1-299) x BD061705 (1-1117)
Qy 1 MetLYsLYsAsnIleLeuAsnLeuAlaLeuValGluAlaLeuSerThrSerPheLeuMet 20
Db 183 ATGAAAGAAATATCTTAATTTAGCTTAGGGCGCGCTTACGCGCTTTTGATG 242
Qy 21 AlaLYsProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLYsLeYsThrThraSpSer 40
Db 243 GCTAAGCCGGCTCATRAACGAGATTAACCTACGATTAACCAACCAAGAAAGAAAGCACTGATTC 302
Qy 41 SerAlaGlyValLeuAlaThrValAspGlyArgProIleThrLYsSerAspPheAspMet 60
Db 303 TCACCCGGGCTGTGAGCAAGTGTAGTGCAGACTATCATCACTAAAGCGAATTTGATG 362
Qy 61 ILeYsGlnArGAsnProAsnPhaSpnPhaSpnLeuYsGlnuYsGlnuA 80
Db 363 ATTAAGCAACGAAATCCATATTGATTTGACAGCTTAAAGAGAAAGAAAGAGCC 422
Qy 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnuAlaLYsThrGluYsLeu 100
Db 423 TTGATTGAGCAAGCTATCCGACCGCAGCTTGTAGAAATGAGGCTTAAGCAAGAAAGCTC 482
Qy 101 AppSerThrProGluPhelYsAlaMetMetGluAlaValYsLeYsGlnAlaLeuValGlu 120
Db 483 GATCAGACTCCAGAAATTTAAAGCGATGATGAGAGCGGTTAAAGAAACAGGCTTTAGTGA 542
Qy 121 PheTPAlaYsLeYsGlnAlaGluGluValYsYsValGlnIleProGluYsGluMet 140
Db 543 TTTTGGGCTAAAGAAACAGGCTGAGAGAGTGAAGAAAGTCCAAATCCAGAAAGAAAG 602
Qy 141 GlnAspPheTYrAsnAlaAsnLYsAspGlnLeuPheValYsGlnuAlaHisAlaArg 160
Db 603 CAAGATTTTTCACACGCTTAATTAAGATCAGCTTTTGTCAAGCAAGAACCCATGCTAG 662
Qy 161 HisIleLeuValYsThrGluAspGluAlaYsArgIleIleSerGluIleAspYsGln 180
Db 663 CATATTTTAGTAAAGAACCGAAGATGAGCTTAACGATTAATTTGAGATTGACAAACAG 722
Qy 181 ProLYsAlaYsLeYsGlnuAlaYsPheIleGluLeuAlaAsnArgAspThrIleAspPro 200
Db 723 CCAAGAGCTTAAAGAAAGAACCCAAATTCATGTAGTTAGCCAAATCGGAGTACGATGATCT 782
Qy 201 AsnSerLYsAsnAlaGlnAsnGlyYsAspLeuGlyYsPheGlnLYsAsnGlnMetAla 220
Db 783 AACACAGAAAGACCGCAAAATGCGGTGATTTGGGAAATTCCAAAGAAACCAATGGCT 842
Qy 221 ProAspPheSerLYsAlaAlaPheAlaLeuThrProGlyAspTYrThrYsThrProVal 240
Db 843 CCGGATTTTCTAAAGCCGCTTCCGCTTACCTTCGGGAGTTAACAATAAAACCCCTGTT 902
Qy 241 LysThrGluPheGlyTYrHisIleIleTYrLeuIleSerLYsAspSerProValThrTYr 260
Db 903 AAAACAGAGTTGGTTATCATATTATCTATTGATTCTTAAGAGTAGCCCTGTAACTTAT 962
Qy 261 ThrTYrGluGlnAlaYsProThrIleYsGlyMetLeuGlnGluYsLeuPheGlnGlu 280

Db 963 ACTTAGACCAAGCTTAACCTTACCTTAAGGGGATGTTACAGAAAGCTTTCCAGAA 1022

Qy 281 ArgMetAsnGlnArgIleGluGluLeuArglybHisAlaIleValIleAsnIle 299

Db 1023 CGCATGAATCAACGCAATTGAGGAATTAGGAGACGCTAAATTTGTTCCACAG 1079

RESULT 8

BD061969

LOCUS BD061969 1082 bp DNA linear PAT 27-AUG-2002

DEFINITION Antigenic composition and method of detection for Helicobacter pylori.

ACCESSION BD061969

VERSION BD061969.1 GI:22607574

KEYWORDS JP 2001517091-A/303.

SOURCE Synthetic construct

ORGANISM other sequences; artificial sequences.

REFERENCE 1 (bases 1 to 1082)

AUTHORS Chow, T. P., Fry, K. E., Lim, M. Y., and Mcatee, C. P.

TITLE Antigenic composition and method of detection for Helicobacter

JOURNAL Patent: JP 2001517091-A 303 02-OCT-2001;

COMMENT GENELABS TECHNOLOGIES INC

PN JP 2001517091-A/303

PD 02-OCT-2001

PF 25-APR-1998 JP 1998547263

PR 25-APR-1997 US 60/045107, 14-OCT-1997 US 60/061958 PI

THREESA P CHOW, KIRK E FRY, MOON Y LIM, C P MCATEE PC

C12N15/31, C07K14/205, C07K16/12, A61K39/106

CC Strandedness: Single;

CC Topology: Linear;

FEATURES

source Location/Qualifiers

1..1082

/organism="synthetic construct"

/mol_type="genomic DNA"

/db_xref="taxon:32630"

ORIGIN

Alignment Scores:

Pred. No.: 4,4e-105 Length: 1082

Score: 1474.00 Matches: 290

Percent Similarity: 97.66% Conservative: 2

Best Local Similarity: 96.99% Mismatches: 7

Query Match: 96.97% Indels: 0

DB: Gaps: 0

US-10-039-183a-2 (1-299) x BD061969 (1-1082)

Qy 1 MetIyGlyAsnIleLeuAsnIleValIleGlyAlaLeuSerThrSerPheLeuMet 20

Db 165 ATGAAATAAATAATCTTAATTAATTAAGGTTAGGGGCGTTGACGCTTTTGAAG 224

Qy 21 AlaIyPProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrIleValSerThrIleAspSer 40

Db 225 GCTTAAGCCGGCTCATACCAAGATTAACGCTTACCAACCAAAAAAGCACTGATTTCT 284

Qy 41 SerAlaGlyValIleuAlaThrValaIlePgiIyArgProIleThrIlySerAspPheAspMet 60

Db 285 TCACCCGGCGGTAGCGACAGTGATGCGACAGCTATACCTAAAGCGATTTTGATATG 344

Qy 61 IleIyGlnIleArgAsnProAsnPheAspPheAspIleuIleGlyIleGlyIleGlyAla 80

Db 345 ATTAAGCAACGAATCTTAATTTTGAACACCTTAAGAGAAAGAAAGAAAGAGCC 404

Qy 81 LeuIleAspGlnAlaIleArgThrAlaLeuValaIleAsnIleValaIleArgThrGlyIleu 100

Db 405 TTGATTGACCAAGCTTACCGACCGCACTTGTAGAAAGAGGCTTAAGCAGAAAGCTC 464

Qy 101 AspSerThrProGluPheIleValaIleMetGlyAlaValaIleIleGlyGlnAlaLeuValGlu 120

Db 465 GATCAGACTCCAGAAATTTAAAGGATGATGAGAAAGCGTTTAAAAAAGAGGCTTTAGTGA 524

Qy 121 PheTrpAlaIleIleGlyGlnAlaIleGluValaIleIleValaIleGlnIleProGluIleGluMet 140

Db 525 TTTGGGCTAAAAAAGAGGCTGAGAGAGTGAAGAAAAAGTCCAAATCCAGAAAAAGAAAT 584

Qy 141 GlnAspPheIleAsnAlaAsnIleAspGlnIlePheValaIleGlnIleValaIleArg 160

Db 585 CAAGATTTTACACCGCTAATTAAGATCAGCTTTTGTCAACAGCAAGAGCCCATGCTAG 644

Qy 161 HisIleLeuValIleThrGluPheGluValaIleArgIleIleSerGluIleAspIleGln 180

Db 645 CATATTGATGATAAAGCCAGAGATGAGGCTAAACGATTAATTTGTAGATGCAAGACAG 704

Qy 181 ProIleAlaIleIleGlyGlnAlaIlePheIleGluLeuAlaAsnArgAspThrIleAspPro 200

Db 705 CCAAGGCTAAAAAAGAGACCAATTCATTGAGTTAGCCAAATCGGATTCGATTCCT 764

Qy 201 AsnSerIleAsnAlaIleGlnIleGlyIleAspLeuGlyIlePheGlnIleAsnGlnMetAla 220

Db 765 AACAGCAAGACGCGCAAAATGCGGTGATTTGGGGAATTCCAAAAGAACCAATAGGCT 824

Qy 221 ProAspPheSerIleAlaIleAlaPheAlaLeuThrProGlyIleAspThrIleThrProVal 240

Db 825 CCGGATTTTTCAAAGCCGCTTCGCTTAACTTCGCGGATTAACCTAAACCCCTGTT 884

Qy 241 LysThrGluPheGlyIleIleIleIleIleIleIleIleIleSerIleAspSerProValIleThr 260

Db 885 AAAACGAGATTTGGTTATATATATATATATTTGATTTCTAAAGATAGCCCTGTAATTTAT 944

Qy 261 ThrIleGluGlnAlaIleAspProThrIleIleGlyIleMetLeuGlnIleIlePheGlnIle 280

Db 945 ACTTAGACCAAGCTTAACCTTACCTTAAGGGGATGTTCAAGAAAGCTTTTCCAGAA 1004

Qy 281 ArgMetAsnGlnArgIleGluGluLeuArglybHisAlaIleValIleAsnIle 299

Db 1005 CGCATGAATCAACGCAATTGAGGAATTAGGAGACGCTAAATTTGTTTCAAGAA 1061

RESULT 9

AX788859

LOCUS AX788859 844 bp DNA linear PAT 17-JUL-2003

DEFINITION Sequence 1323 from Patent WO2006501.

ACCESSION AX788859

VERSION AX788859.1 GI:32955273

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1

AUTHORS Legrain, P., Rain, J. C., Colland, F., de Reuse, H. and Labigne, A.

TITLE Protein-protein interactions in Helicobacter pylori

JOURNAL Patent: WO 0206501-A 1323 29-AUG-2002;

Hybridgenics (FR) ; INSTITUT PASTEUR (FR)

FEATURES

source Location/Qualifiers

1..844

/organism="Helicobacter pylori"

/mol_type="unassigned DNA"

/db_xref="taxon:210"

ORIGIN

Alignment Scores:

Pred. No.: 1,32e-92 Length: 844

Score: 1311.00 Matches: 257

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 86.25% Indels: 0

DB: Gaps: 0

US-10-039-183a-2 (1-299) x AX788859 (1-844)

Qy 43 GlyValIleuAlaIleThrValaIlePgiIyArgProIleThrIlySerAspPheAspMetIleIle 62

Db 3 GCGGTGTTAGCAGAGTGATGCGACAGCTTACCTTAAGAGGATTTTGAATGATTAAG 62

QY 63 GlnArgAsnProAsnPheAspPheAspLysLeuLysGluLysGluLysGluLysLeuLeu 82
DB 63 CAACGAAGATCCCAATTTTGGTATTTTGGCAAGCTTTAAAGAGAAAAGAAAAGAACCTTTGATT 122
QY 83 AspGlnAlaIleArgThrAlaLeuValGluAsnGluAlaLysThrGluLysLeuAspSer 102
DB 123 GATCAAGCATATTCGACCCGCTTGAGAAAATGAAAGCTTAAACCGAGAAATTTGGACAGC 182
QY 103 ThrProGluPheLysAlaMetMetGluAlaValLysLysGluAlaLeuValGluPheTrp 122
DB 183 ACTCCAGAAATTTAAAGCGATGATGAGAGCGGTTTAAAGAACAGGCTTTAGTGAATTTTGG 242
QY 123 AlaLysLysGluAlaGluLysValLysLysValGlnIleProGluLysGluMetGlnAsp 142
DB 243 GCTAAAGAAACGCGCTGAAGAGTGAAGAAAAGTCCAAATCCGAGAAAAGAAAATGCAAGAT 302
QY 143 PheTYrAsnAlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHisIle 162
DB 303 TTTTCAACGCTTAACAAAGATCAGCTTTTGTCAAGCAAGAACCCCATCTAGCGCATATT 362
QY 163 LeuValLysThrGluAspGluAlaLysArgIleIleSerGluIleAspLysGlnProLys 182
DB 363 TTAGTGAAGAAACCGAAGATGAGCTTAAACGGAATTTTGTGAGATTGACAAACAGCCAAAG 422
QY 183 AlaLysLysGluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsnSer 202
DB 423 GCTAAAGAAAGAGCTTAATTCATTGATTGAGCCAAATCCGGATACGATTCCTTAACAGC 482
QY 203 LysAsnAlaGlnAsnGlyGlyAspLeuGlyLysPheGlnLysAsnGlnMetAlaProAsp 222
DB 483 AAGAACGCGCAAAATGCGCGTATTTGGGAAATTCCAAAAGAACCAATGCTCCGCGAT 542
QY 223 PheSerLysAlaAlaPheAlaLeuThrProGlyAspTYrThrLysThrProValLysThr 242
DB 543 TTTCTAAAGCGCTTTCGCTTAACTCTCGAGGATTTACCTTAACCCCTGTAAAGAA 602
QY 243 GluPheGlyTYrHisIleIleTYrLeuIleSerLysAspSerProValThrTYrThrTYr 262
DB 603 GAGTTTGATTATCATATTATCTATTGATTCTTAAGATAGCCCTGTACTTATCTTAT 662
QY 263 GluGlnAlaLysProThrIleLysGlyMetLeuGlnGluLysLeuPheGlnGluArgMet 282
DB 663 GAACAGCGCTAAACCTTACCTTAAGGAGATGTTTACAAAGAAAGCTTTCCAAAGACGATG 722
QY 283 AsnGlnArgIleGluLysLeuArgLysHisAlaLysIleValIleAsnLys 299
DB 723 AATCAACGCAATGAGAACTAAGAAAGCAAGCTTAATTTTATCAACAG 773

RESULT 10
BD061671 616 bp DNA linear PAT 27-AUG-2002
LOCUS Antigenic composition and method of detection for Helicobacter pylori.
DEFINITION

ACCESSION BD061671
VERSION BD061671.1 GI:22607276
KEYWORDS JP 2001517091-A/5.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 616)
AUTHORS Chow,T.P., Fry,K.E., Lim,M.Y. and Mcatee,C.P.
TITLE Antigenic composition and method of detection for Helicobacter pylori.
JOURNAL Patent: JP 2001517091-A 5 02-OCT-2001;
GENELABS TECHNOLOGIES INC
COMMENT
PN JP 2001517091-A/5
PD 02-OCT-2001
PR 25-APR-1998 JP 1998547263
THREESA P CHOW,KIRX B FRY,MOON Y LIM C P MCATEE PC
C12N15/31,C07K14/205,C07K16/12,A61K39/106
CC Strandedness: Single;
Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
source Location/Qualifiers
1..616
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN

Alignment Scores:

Pred. No.:	4,92e-70	Length:	616
Score:	1017.00	Matches:	200
Percent Similarity:	98.53%	Conservative:	1
Best Local Similarity:	98.04%	Mismatches:	3
Query Match:	66.91%	Indels:	0
DB:	6	Gaps:	0

US-10-039-183a-2 (1-299) x BD061671 (1-616)

QY 73 LeuLysGluLysGluLysGluLysGluLysLeuIleAspGlnAlaIleArgThrAlaLeuValGlu 92
DB 3 CTTAAAGAGAAAGAAAAGAGCTTGATTTGAGCAAGCTATCCGACCGCAGCTTGTAGAA 62
QY 93 AsnGluAlaLysThrGluLysLeuAspSerThrProGluPheLysAlaMetMetGluAla 112
DB 63 AATGAGGCTTAAGCGCAAGAAAGCTCGATCACTCCAGAAATTTAAAGCGATGAGAGCG 122
QY 113 ValLysLysGlnAlaLeuValGluPheTrpAlaLysLysGlnAlaGluLysValLysLys 132
DB 123 GTTAAAGAAACAGGCTTTTACTGTGAATTTTGGCTTAAAGAACAGCGTGAAGAGTGAAGAAA 182
QY 133 ValGlnIleProGluLysGluMetGlnAspPheTYrAsnAlaAsnLysAspGlnLeuPhe 152
DB 183 GTCCAAATCCCGAAGAAAGAAATGCAAGATTTTTCACAGCTTAATTAAGATCAGCTTTT 242
QY 153 ValLysGlnGluAlaHisAlaArgHisIleLeuValLysThrGluAspGluAlaLysArg 172
DB 243 GTCAGCAAGAAAGCCATGCTAGCATTTTAAAGAAAACGAGATGAGCTTAAACGG 302
QY 173 IleIleSerGluIleAspLysGlnProLysAlaLysLysGluAlaLysPheIleGluLeu 192
DB 303 ATTATTTCTGAGATTGACAAACAGCCAAAGCTTAAAGAAAGAACCAATTCATTGAGTTA 362
QY 193 AlaAsnArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyLysAspLeuGly 212
DB 363 GCCAATCCGGATATGATGATCTTAACAGCAAGAACGCGCAAAATGCGCGTATTTGGCG 422
QY 213 LysPheGlnLysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeuThrPro 232
DB 423 AATTCCAAGAAAGAACCAAAATGCGCTCGGATTTTCTAAAGCCGCTTTCCTTAACCTCT 482
QY 233 GlyAspTYrThrLysThrProValLysThrGluPheGlyTYrHisIleIleTYrLeuIle 252
DB 483 GGGGATTAACCTTAAGAACCCCTGTAAACAGAGTTTGGTATCATATTATCTTATTGATT 542
QY 253 SerLysAspSerProValThrTYrThrTYrGluGlnAlaLysProThrIleLysGlyMet 272
DB 543 TCTAAAGATAGCCCTGTAACTTATATGAGCAAGCTAAACCTTAAGGGGATG 602
QY 273 LeuGlnGluLys 276
DB 603 TTCAAGAGAAAG 614

RESULT 11
AX790177 595 bp DNA linear PAT 17-JUL-2003
LOCUS Sequence 2641 from Patent WO0206501.
DEFINITION AX790177
ACCESSION AX790177
VERSION AX790177.1 GI:32955932
KEYWORDS
SOURCE Helicobacter pylori
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE 1

AUTHORS Legrain, P., Rain, J.C., Colland, F., de Reuse, H. and Labigne, A.
TITLE Protein-protein interactions in *Helicobacter pylori*
JOURNAL Patient: WO 0206501-A 2641 29-AUG-2002;
Hybridgenics (FR) ; INSTITUT PASTEUR (FR)
FEATURES
source 1. .595
/organism="Helicobacter pylori"
/mol_type="unassigned DNA"
/db_xref="taxon:210"
ORIGIN
Alignment Scores:
Pred. No.: 8,99e-43 length: 595
Score: 664.00 Matches: 129
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.68% Indels: 0
DB: Gaps: 0
US-10-039-183a-2 (1-299) x AX790177 (1-595)
QY 171 LysaTglllellesetGulleaplyeGlnProlysaLalysylGluAlalyshelle 190
DB 3 AAACGATTATTTCGAGATTGACAAACAGCCAAAGCTAAAGAAAGAACTAAATTCATT 62
QY 191 GllleuAlaAaanaTgaBpThrlleapProaanserLyasaAlaGlnaenGlyglYaap 210
DB 63 GAGTAGCCAAATCGGATCGATGATGCTTAACAGCAAGACCGCAAAATGCGGTGAT 122
QY 211 LeuGlLysePhheGlnlysaenGlnMeAlaProaspPheSerLyasaAlaAalPheAlaLeu 230
DB 123 TTGGGGAATTCACAAAGAACCAATGCGTCCGAGATTTCTTAAGCCGCTTCGTTTA 182
QY 231 ThrProGlyaBpTyThrllyeThrProAllyeThGlnPhheGlyTyThsileleTy 250
DB 183 ACTCGTGGGATTAACACTAAACCCCTGTTAAACAGAGCTTGATCATATTATCATAT 242
QY 251 LeuileSerLyasBpSerProAllyeThrTyThrTyThGlnAlalyshProthrlleTy 270
DB 243 TTGATTTCCTAAAGATAGCCCTGTAACCTTAATTAAGCAAGCTTAACCTTAAG 302
QY 271 GlyMetLeuGlnGluLyseupPhheGlnGluArGmeArasGlnArglleglGluLeuAaX 290
DB 303 GGGATGTTCAAGAAAAGCTTTTCCAGAACCGATATCAACGCAATTGAGAACTTAAGA 362
QY 291 LyshlAalalyeIleValilleaenly 299
DB 363 AAGCAGCTAAATGTTATCAACAG 389
RESULT 12
AE017144 303820 bp DNA linear BCT 26-JUN-2003
LOCUS Helicobacter hepaticus ATCC 51449 section 1 of 6 of the complete
DEFINITION genome.
ACCESSION AE017144 AE017125
VERSION AE017144.1 GI:32261547
KEYWORDS
SOURCE Helicobacter hepaticus ATCC 51449
ORGANISM Helicobacter hepaticus ATCC 51449
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE
1 (bases 1 to 303820)
Suerbaum, S., Josenhans, C., Sterzenbach, T., Drescher, B., Brandt, P.,
Bell, M., Droge, M., Fairmann, B., Fischer, H. P., Ge, Z., Horster, A.,
Holland, R., Klein, K., Konig, J., Macko, L., Mendz, G. L., Nyakatura, G.,
Schauer, D. B., Shen, Z., Weber, J., Frosch, M. and Fox, J. G.
The complete genome sequence of the carcinogenic bacterium
Helicobacter hepaticus
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7901-7906 (2003)
JOURNAL MEDLINE
PUBMED 12810954
REFERENCE 2 (bases 1 to 303820)
Drescher, B. and Suerbaum, S.

TITLE Direct Submission
JOURNAL Submitted (27-MAY-2003) Institute of Hygiene and Microbiology,
University of Wuerzburg, Josef Schneider Str. 2, Wuerzburg D-97080,
Germany
FEATURES
source 1. .303820
/organism="Helicobacter hepaticus ATCC 51449"
/mol_type="genomic DNA"
/strain="ATCC 51449"
/db_xref="ATCC:51449"
/db_xref="taxon:235279"
141. .2423
141. .2423
/locus_tag="HH0001"
141. .2423
/locus_tag="HH0001"
/note="similar to Helicobacter pylori toxin-like outer
membrane protein"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAP76598.1"
/db_xref="GI:32261548"
/translation="MONTNIYQKTFEVLFLSMAPFSKQNTKRLPSSILLONGE
SRSDILTSVGNQYCAFGIMSNMAGSYDFSSTHNGSSRLSVATPTPVRHRSVF
YRHLNVGENSTLSINDPHTFALSGGLTNNAALEIILQKEKI PASEFAPSSKVH
SRNAPVYNGSTLNIISNTDFVILSSKGI INNAQMIIDSSTIRPONTLHNGTELT
GIVNATGOSQTITNGTSHFTINYNKNGNRYNGTSPADSVQSSWMQNDAPRGGC
NLINYGTHITLGNLINOVELGKTQNSVOIYGGTILVDGKNGKANTTLICMTN
GKMGQIDGNTVNRGIAKVIDTATLNTDHRINGWHTGLNANRDVLNGSSSEFLNA
RISHQNTTITLNRKQINAFNMWTLPIQSTLITLLENSLYNPDTLQOQSIYAAGN
KNYTRKLNINIESASMLYASPLNIMAVONSIMPLAQVTPPRKSVLINIAPLS
IASRGSLSPLNGVNIISAOI PYKTHLSAFAYGMNATHTLTLQTHEDTNSFLVGL
YDRISLPYEVSLIAYGASFKARKHRLINNSAOTRYSVHEIGLSAQIHLAPREK
LWKPFQINVAFLQSAFKENLVSDTTSAMLTAPMWSHLQLSICTGQYLYKKQY
ELFGAQIQVALTNSFSFAHFGSSVLNFPQNSLGLTLFPGSMPLAKFSLSAYTFY
SRSHLAFOSYSTIMLSYFF"
2484. .3989
/gene="lrpC"
/locus_tag="HH0002"
2484. .3989
/gene="lrpC"
/locus_tag="HH0002"
/EC_number="5.3.1.24"
/codon_start=1
/transl_table=11
/product="phosphoribosylanthranilate isomerase"
/protein_id="AAP76599.1"
/db_xref="GI:32261549"
/translation="MVEVLAKITTORSDITLKGFPGEYEPKORHLPLIKPLDSVF
FLAEIKRSSPSNGTKETPSYTKLADYINGVGAISVCEHYGSGSLDMSYKSA
YFNACILRKDPFIQYOEVDVSYRAGADNVLIVANFINDGFSHPKMIYEECLAKYGI
TPLEIVHTQAEIDTAPLOAPLIGINSRLHFEINIPACALKNALSSKILIFESGI
DSYSAFPIGSLDPGLCGSYLVSHKQPTALQSLKPAFTLGTQKPRFVYIPESL
SNHSPQSKFLKICGITNLDDALMWGOEKIDMLGFLIVGHSPTYESKQIKDIKALQ
KLYPHILRAVINDDKOALNEAKSLYQOGYIDAIQHALNPLPDI FAKSDLEALFC
PYPNONIAHIDPFTPYEGIFCLYDSKSAQGGSGSINKDVLRSLOERYLTIAGSIN
PHNIADFLAKRTLLDINSIGIEKVGKDIHKLRLDLNLRNAISSTKEQK"
3986. .5032
/locus_tag="HH0003"
3986. .5032
/locus_tag="HH0003"
/note="similar to Streptomyces coelicolor putative
lipoprotein"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAP76600.1"
/db_xref="GI:32261550"
/translation="MKILVTLIFSLMSICVATAKELNLIAGDALLSHSTYVKAKO
EDGSYDFSSMLTALKPVEEKDLATYNDETILGVKLGSLTFNFSPOFSDNMLKYA
GNLISLANHNTLDGRGAVRSLSLAYWEKPVLTAGSYNSFLERTPKLEKNGIKYA
LLAYVTNGNIGLIPKGEYLVVVYTKAMEKDIKARSQVDFLWYSMHMGIEDFPS
KEQDLAKFLADLGVDLIIGTHPHVVOQPAEMVIGNLTLYVSLGNLISGQGTNKRIGML

Db 104547 AAGCAAGGACGAGGAAATTCATAGAACTTGCAATATGCAAAATCGATTGACCCAGCT 104606
CDS
Qy 202 SerIysenAlaIGlaSngIyGlyAspLeuGIyIyPheGlnIyLysaGlnIleAlaPro 221
Db 104607 TCAAAACAACAAAATGCTGGGATTTAGCGATTTAAACGTCGGGATAGATCTT 104666
Qy 222 AspSerIySaIaAlaPheAlaLeuThrProGIyAspTYrThrIyThrProValIyS 241
Db 104667 ATGTTTCAAGCTGCTTTTACTTAAACCGGACTTACACAAAAGAGCTGTCTT 104726
Qy 242 ThrGIuPheGIyTYrHisIleIeTYrIleuIleSerIyAspSerProValIyThr 261
Db 104727 ACGCAATTTGGCTATCATATTTATCTTGAACGCAAAAGTAGGCTTAAGCTCATTTCCC 104786
Qy 262 TYrGIuGlnAlaIyProThrIleIySgIyMeIleGlnGlnIyLeuPheGlnIyGly 281
Db 104787 TATTAAGATGCTAAATAAAATTTATGAAATAGCATCAAAATGCAAGTATTCAGAGTGT 104846
Qy 282 MetAaGlnArgIleGIuGlnIyLeuArgIySHIaIaIyIleValIleAaIyS 299
Db 104847 ATGATGCAAAAGATTCAGGCTTTACCGCTTAAGCTAAATCAAAATTCAGAAA 104900
RESULT 13
LOCUS BX571660 349926 bp DNA linear BCT 08-SEP-2003
DEFINITION Wolinella succinogenes, complete genome; segment 4/7.
ACCESSION BX571660 BX571656
VERSION BX571660.1 GI:34483186
KEYWORDS complete genome.
SOURCE Wolinella succinogenes
ORGANISM Wolinella succinogenes
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
Helicobacteraceae; Wolinella.
REFERENCE
AUTHORS 1
Schuster, S.C., Baar, C., Eppinger, M., Raddatz, G., Simon, J.M.,
Lanz, C., Kilmek, O., Nandakumar, R., Gross, R., Rosinus, A.,
Keller, H., Jagtap, P., Linke, B., Meyer, F. and Lederer, H.
TITLE Complete genome sequence and analysis of Wolinella succinogenes
JOURNAL Unpublished
2
Schuster, S.C.
Direct Submission
Submitted (15-MAY-2003) Max-Planck Institut for Developmental
Biology, Spemannstr. 35, 72076 Tuebingen, GERMANY
FEATURES
SOURCE
1. 349926
/organism="Wolinella succinogenes"
/mol_type="genomic DNA"
/strain="DSMZ 1740"
/db_xref="taxon:844"
complement(716..2527)
/locus_tag="WS1091"
complement(716..2527)
/locus_tag="WS1091"
/note="CONTIGUAL TRANSFER PROTEIN TRAG"
/codon_start=1
/transl_table=11
/product="CONTIGUAL TRANSFER PROTEIN (TRAG)"
/protein_id="CAE10185.1"
/db_xref="GI:34483187"
/translation="MIFQKGNHPIIVIFGMSLLLIYSIRAFVSSKRRFELATKDI
LKALPKRLSPVLALALPLPAISLSTFPEPLDTPKIKOKLIESIDSNISLKN
EHWLVANALTPSLALALYLPFAKSTAYGRKRMAREKDIQKGLNCEGRFLAARQ
KIKYDTPSLTVIAPPGTGKTSAILIPULSIPNSOVIVDVGESICSLTAQYRKUL
LNQVIFPNPGRDSNFHNPFLERMSKGLFAEKTAIVQVSGTIFPKPKDEKSHR
EIAKTFPEFAAHNIKEYTTLYEYRPKDWTDELDEKDFDEKKEEAGEEVT
PALLRQIAENETLNDNRDARFLGTAFANEGSVLSTFKMSIFGDLRYKELETD
NSFMRRLAENDITLYIKLEKIDPSLPIIILLETYKEIMSRSESDPKRYILEI
DEVARFGEIDPEIIEPISRSINLNIPAGSYAQIRKYSOEDIEIMDTAYQVVF
RNRQAAEDISKEVGFPAHKKTTSQDKILQSTISEEKRLVTADIDINIPDK
VLIAHGKATYVLAQANFYFDRTERRQTKIKENK"
complement(2880..3791)
/gene="VIRB11"

CDS
/locus_tag="WS1092"
complement(2880..3791)
/gene="VIRB11"
/locus_tag="WS1092"
/codon_start=1
/transl_table=11
/product="VIRB11"
/protein_id="CAE10186.1"
/db_xref="GI:34483188"
/translation="WTYVQLNLTLEBYLTADGVNRYVIRREVMIDKGGWHPRI
FDARITFLKPCIEIENSRMRENPFCSCAIPQTRRVQLHSYITLDSIAL
SIRVPSNIFPEIENFLISPNIKTHDILKRLISERKNVLSGTSKGSFLNALRL
VDRODRIVITIEDEBELFVENPNRQIOLVPKLIDYKMSYELANSMRLRPDILLGEI
DTKMTYLRSLNTHSGMSLSTLHANSPADAIHAIINAMRSGQVVASIMRGYIKSA
IDVLIQIRBEKSRRIIDILNLEKALS"
complement(3967..6027)
/locus_tag="WS1093"
complement(3967..6027)
/locus_tag="WS1093"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE10187.1"
/db_xref="GI:34483189"
/translation="MRQIMLFVLSVCLFGASKDNSSGDKLDVTIKKDTILKYQMI
EAKPLVNDBELILRGYKTAPEGKVEINPQAEYQDQTYNLAPEKSAFLKDSNA
TLKNSKTIIVGGSKERELATIKTKLEKEREKESNAGTGAATASGSGSNGS
NGITSVNPYGSIDSTSKDSTNSRVSCEPSAFVQGIATVYIEENQCTPYOSNNIYE
KYNATCKNRIDYANNTIELGVELTTIDGKEARVSTCTWKEALIPLESVGYCATP
DYDNGVALIQEYVYLLDEKKTIGSCPSGSKLELKHDTNSCTDRHDFINGITYDQ
TOYFYSYNNTRHNVECVNDPDLPYMHVEDTTCMQQLEGRVFKYRLAYKDMNH
VRVATRECTITSGLELIEEPAGYLYMDSHQALRKIKYFIVPNTKRIYDKQVOT
SSYVREEDCGMDNDPMRVSYSKRPVPTDNNRVDRNEKTKTQISVYVTD
KATTTTLETLNKTTPDGTGKTYGAEAYISNTNRYRVVSLDSARTIPHPFPPS
YSWNSCPQWYTEGTQFSPSDTIKVAQNYKMERHSTYNLIKQDGSCTYITSL
DPNNTYGTGTYRVAIQNSPHANNQTIIDIVQTESKHXYIRDDGSTYEDPESWQETR
YIAK"
complement(6024..7127)
/gene="COMB3"
/locus_tag="WS1094"
complement(6024..7127)
/gene="COMB3"
/locus_tag="WS1094"
/codon_start=1
/transl_table=11
/product="COMB3"
/protein_id="CAE10188.1"
/db_xref="GI:34483190"
/translation="MKPVKAILISSVVLILMGSFVAVMDSKEDSSDEGFIDET
KPLGYLVKESPKERTITTEKKEPEKDTLAEISTLETTIKETTKTATYMDIIFPRDQ
EARLKEQROSLAORLGAPIOINPINOFPQEQEIDYGNARPSNYKODLASMEHR
LRLTTADRMMPAILITPISDLPGQVALLEBDIFAMGRANLVPKSRALGIGSN
NKGLENRLQIVWRIITPQGVNIIITQADVADSTMSGALGLDKKYVDRGLASLS
TVANAILITVANIQOTKFPNDHOTTLTLDKSGNDSTIMNNIIEQOTKLAIPITISAGS
RIFINPTVDIWPIDKNEVLVQYTFEYKEFEK"
complement(7124..8053)
/gene="COMB2"
/locus_tag="WS1095"
complement(7124..8053)
/gene="COMB2"
/locus_tag="WS1095"
/codon_start=1
/transl_table=11
/product="COMB2"
/protein_id="CAE10189.1"
/db_xref="GI:34483191"
/translation="MMKNIWLLSLASALMAEVSIFDEGEMSTKPPSPMATPMTID
IKATQOIRLKFSGENENIISINPHGCTPPIRTRYAMSTTIILBEDRIIDILGDSV
GPEKTIPEKDKYDLSNITIIYVKMGICITSLTIGESRLKSVLYXSDHNSKTXPNL
VVFSEKRDQITRIIRNLEKAEFLAKAEKAEKAKKISLEADYILIDQEVNRLN
VKIIEIERGVQDEGDSLKAVDIFNDNKFYFRYDRNRADSRFPVYRVVDDYDNPVN
TRVVGNYIIATISNRFILRAGEKVVCVKQKK"
complement(8050..8694)
/gene="VIRB11"

gene
CDS
/gene="COMB1"
/locus_tag="WS1096"
complement(8050..8694)
/gene="COMB1"
/locus_tag="WS1096"
/codon_start=1
/transl_table=11
/product="COMB1"
/protein_id="CAE10190.1"
/db_xref="GI:34483192"
/translation="MEKRVATSLHRLKRLIDVLLAIACTAEILVIVLTITLPL
PLKEKPYLVHFSNABQNFVTLERAGAEIRSKAVLTALMTALVYKRRKIDMTKER
FEVERLOQSAKWRSFETLVNTENS1YAKKLRKRLVNISFLVAGISTVDFIKKT
NGENSGTISEERYRAVLISYGFADTKLRLEKKNMPTGLVNDYGVTLQINIGDAL"
complement(8842..9843)
/locus_tag="WS1098"
complement(8842..9843)
/locus_tag="WS1098"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE10191.1"
/db_xref="GI:34483193"
/translation="MKMATGBGIGAYHYLSIMIGAFOENVVNNMLEAQAQIYNNSWP
EASFLVILITVAFQWQGGTREGVLRILNILLMA1LVKAILIWNELPFLGVNQLP
ADIFASAVSFVVGELSESPH1ILDQLTVLGDVGVLIDNMGTVSNVFPALFAMI
FVLVSLAVLVIIFMLVSKMSIVALGLAPILPLFKRGMFGMKRYISYL
YGPFTIFPGYLAISTLIAIOTKADMTSPDLGNLPAIIAIVIIIGIVYWKIPOVNO
IIGTQNDGSLVGAINGTIAASYASRAATQGMQVQKARSDMLAKKHSKIKTSS"
complement(9840..12128)
/locus_tag="WS1099"
complement(9840..12128)
/locus_tag="WS1099"
/codon_start=1
/transl_table=11
/product="VIRB4 HOMOLOG"
/protein_id="CAE10192.1"
/db_xref="GI:34483194"
/translation="WNILRPLLELISKPMHSIAEDNNIIGIYDSAGIIVTKDNVA
IGFEIGLSYGGLTPEDEALHLTRKIPKAKNNPIRIDISYKKSLSLHHRARAKP
HSSRIQRRBNFNARLSYIILSTKKGILSPLEBRKRTTSGENSLKIKIOLK
DAAKMLYLGRFRPALDGBELMVAYACMSPLCSYKQGYLRDSTIDSNILFKR
DYI1HERDSEISRYIAIKAYDTEI1SS1ISTFEN1PVSNNVFOI051SKSEVW
KSKQK1KMSQNELVVAELDTLKEIQSDREMLLYFSFAMVSATQOEBLQRCHEV
KALVAKGLIARETINOQPLYSFPGRTINSTRLOSSTALISTLFEKDIEGSSN
SMGSPVTKFKNITETPYEPFNHMSBGDCPMSHTLVGNVYAGTTMSPLMTCLTK
PEIDI1PALKRAGMNPCTFLGDDYSDINDERLNPISLAPTHENS1PLNSTRILGN
IDEREYEVNATADTLERTYENSENTTLIDFYESLERVPNLKRYKSHUDGI.FAHPD
ALSFOKQITILNMDSVLKDDEIKASHLAYVFKILYQAKTKGKAFFVDELKDYLAN
DVMAHKLETLLEIRKNGVALGVONLDFFDNIKNASGMSMAHFA1FPTTNESI
OGHIELTPALELFLKNTNPLERKLMLKNAKTESYILDVNLRLREYLAVFSSAEDV
LQMKALABEGEMREKYLQREG"
complement(12112..12357)
/locus_tag="WS1100"
complement(12112..12357)
/locus_tag="WS1100"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE10193.1"
/db_xref="GI:34483195"
/translation="MIEANNIREITKKEKIKGLSRATWALSMI1AGGSFPHILFWNIP
FFIPLCFPVYVEFPDEDMDIFSAKIELKAKNEYFA"
complement(12366..12764)
/locus_tag="WS1101"
complement(12366..12764)
/locus_tag="WS1101"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE10194.1"
/db_xref="GI:34483196"
/translation="WRSNHPPLPLSTRLKGSTIGRISPPSPISSTALPTHONKQMK

QIWMILSLPBLAASGDPGAKFAEIVERYDMIQNGVAVASLVIIIMGFMLFKM
FEOFKMVFVGIIGIALYGGTFIVNALVN"
13628..16075
/locus_tag="WS1102"
13628..16075
CDS
Alignment Scores:
Pred. No.: 9,996-26 Length: 349926
Score: 482.50 Matches: 101
Percent Similarity: 60.78% Conservative: 54
Best Local Similarity: 39.61% Mismatches: 93
Query Match: 31.74% Indels: 7
DB: 1 Gaps: 3
US-10-039-183a-2 (1-299) x BX571660 (1-349926)
QY 41 SerIlaGlyValLeuAlaThrValAspGlyArgProIleThrLysSerAspMet 60
DB 191953 AGTGCTTAAACATTGGCAAGCGCTCGATGAGATCACCAGATTAAGACATTAGCGTC 191894
QY 61 IleLysGlnArgAsnProAsnPheAspLysLeuLysGluLysGluLysGluAla 80
DB 191893 ATGCTTCAGACCATGCTCGAGTGTGATGATCACTTCTCGAAGATATGCAAAAGAA 191834
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGluAlaLysThrGluLysLeu 100
DB 191833 GCTTATGAGCAAGCATTTAGCGGCAAAATGCTTGAACCAAGCAAGGAGGAGAT 191774
QY 101 AspSerThrProGluPheLysAlaMetMetGluAlaValLysLysGlnAlaLeuValGlu 120
DB 191773 CAGAACTCCAGAGACTTAAAGACCGCTTACGATGCGMAAGAGATTGACGCTTAA 191714
QY 121 PheThrAlaLysLysGlnAlaGluGluValLysLysValGlnIleProGluLysGluMet 140
DB 191713 GTTTGGATGCGACGACATGATTAATCTTAAG-----GTCCTGAAGCGCATATG 191663
QY 141 GlnAspPheTyAsnAlaAsnLysAspGlnLeuPheValLysGlnGluAlaIleAlaArg 160
DB 191662 CGAAATATTTATGATGAGAAATTAAGAAAGATTGTTACGCTGAGCTGTAAGCGCAAG 191603
QY 161 HisIleLeuValLysThrGluAspGluAlaLysArgIleIleSerGluIleAspLysGln 180
DB 191602 CATATTTGGTGCAAAACGAGAAAGAGGCGAAAGGTGATGCCGAAATTTGGCAAA--- 191546
QY 181 ProLysAlaLysLysGluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspPro 200
DB 191545 GCAAGTCCCAAGCGCAGGAGGAAGATTAGCAGCTTCTTAAGAGCAAAATCCATGACCT 191486
QY 201 AsnSerLysAsnAlaGlnAsnGlyLysAspLeuGlyLysPheGlnLysAsnGlnMetAla 220
DB 191485 -----GCGGGGCAAAATGAGGAGGAGCTTGATGATTAAAGATCAGATGGTG 191435
QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTyThrLysPheProVal 240
DB 191434 CTTGAGTTTGCACACGACGCTTGGCTTCAAAAAGAGATTAAGCAAGACCCCTGTG 191375
QY 241 LysThrGluPheGlyTyHisIleIleLeuTyLeuIleSerLysAspSerProValThrTy 260
DB 191374 AAGACTCAATTTGGTACCACTTATCTATCTGTAAGATTAATAAAGCCAGCGCTACTC 191315
QY 261 ThrTyGluGlnAlaLysProThrIleLysGlyMetLeuGlnGlnLysLeuPheGlnGlu 280
DB 191314 CCGTATGAGGATGTCAACCTCAATTCAGCAGATCTTAAATCAAAAATTCAGAGAC 191255
QY 281 ArgMetAsnGlnArgIleGluGluLeuArgLysHisAlaLysIle 295
DB 191254 AGCGTCTCTCTACAGCAAAAAGCTTGAAGAGAGGCTCAAGTC 191210
RESULT 14
CUGBF2 CUGBF2 2006 bp DNA linear BCT 05-MAR-1996
LOCUS C.jejuni cbf2 and fdac genes.
DEFINITION X84703
ACCESSION

N-(5'-phosphoribosyl)anthranilate isomerase (199 aa) (EC 5.3.1.24), fasta scores; opt: 271 z-score: 352.4 E(): 2.3e-12, 31.0% identity in 200 aa overlap, and to the trpC domain of many trpC proteins e.g. TRPC_SALT1
 indole-3-glycerol phosphate synthase, fasta scores; opt: 307 z-score: 394.5 E(): 1e-14, 30.8% identity in 201 aa overlap, 38.2% identity to C-terminus of HPI179 (trpC) contains Pfam match to entry PF00697 PRAI, N-(5'-phosphoribosyl)anthranilate (PRA) isomerase, score 102.30, E-value 9.5e-27"
 /codon_start=1
 /transl_table=1
 /product="N-(5'-phosphoribosyl)anthranilate isomerase"
 /protein_id="CAB74184.1"
 /db_xref="GI:6967622"
 /db_xref="GDA:Q9P1F3"
 /db_xref="UniProt/Swiss-Prot:Q9P1F3"
 /translation="MLKIKICIKDEKNAKDLAFNIDFGLIFAKSPRRVSLQQRANLSAIFHEKDKKVVGVFVNDLEQLQIKIKKALDGIQIRTTKEPEILKQNVFWOVISVENSILDKSEIFANLVLEDKGIKGGNGISPDWTLGSTRKDFILAGIGLDNVHAKVTKAKIIDLNSKLEDEKGLDKIKIKIQLKELKK"
 4143..4721
 /gene="trpB"
 /note="Pfam match to entry PF00697 PRAI, N-(5'-phosphoribosyl)anthranilate (PRA) isomerase, score 102.30, E-value 9.5e-27"
 4730..5908
 /gene="trpB"
 /note="trpB"
 4730..5908
 /gene="trpB"
 /note="trpB"
 /EC_number="4.2.1.20"
 /note="Cj0348. trpB, probable tryptophan synthase beta chain, len: 392 aa; highly similar to many e.g. TRPB_THEMA tryptophan synthase beta chain (EC 4.2.1.20) (389 aa), fasta scores; opt: 1476 z-score: 2082.2 E(): 0, 59.9% identity in 384 aa overlap, 56.4% identity to HPI1278. Conatacins P800168 Tryptophan synthase beta chain pyridoxal-phosphate attachment site, P800017 ATP/GTP-binding site motif A (P-loop), and Pfam match to entry PF00247 trp_synB, tryptophan synthase, beta chain, score 806.40, E-value 1e-238"
 /codon_start=1

US-10-039-183a-2 (1-299) x CJ11168X2 (1-308601)

Percent Similarity: 1.35e-17 Length: 308601
 Best Local Similarity: 52.38% Matches: 98
 Query Match: 33.33% Conservative: 56
 24.77% Mismatches: 113
 Indels: 27
 Gaps: 5

7 AenLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMetAlaLysProAlaHisAsn 26
 239127 AACCTTAATGCAAAAATTTCTTAGTCGACCACTTGTATGCGAGTGTATTAAT 239186
 27 AlaAsnAsnAlaThrHisAsnThrLysLysTrpThrAspSerSerAlaGlyValLeuAla 46
 239187 GTAAATGCGAGTACA-----GTAAGCT 239207
 47 ThrValAspGlyLysGProIleThrLysSerAspPheAsp-----MetIle 61
 239208 ACTGTTAATGCGAAGCATTTAGCGATACAGAGTAAGTAATTTTGGCCCTATGCTT 239267
 62 LysGlnArgAsnProAsnPheAspPheAspLysLeuLysGlnLysGlnLysGlnLys 81
 239268 AGAGGACAGC-----GATTTTAAACCTTGTCCAGATATCAAAAAAGCTCTT 239315
 82 IleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGlnLysLeuAsp 101
 239316 ATTCAGCAATATATATATGCAAGATTTAATTTTGCAAGATGCTAAAAACAATTTAGAA 239375

102 SerThrProGluPheLysAlaMetMetGluAlaValLysLysGlnAlaLeuValGluPhe 121
 239376 AAAGACCTTTATACAAAAAGAACTGATGTCGAAAAAGATGCAATACTGTTATGTT 239435
 122 TrpAlaLysLysGlnAlaGluGluValLysLysValGlnIleProGluLysGluMetGln 141
 239436 TATCAAGAGAAAATTTTAAATCTACTTTTAA-----ATTGATGCGGCTAAAGTTAA 239486
 142 AspPheTyranAlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHis 161
 239487 GCTTTTATGATCAAAATTAAGACAAATATGTAACCTCGCAAGATGCGACAAACAT 239546
 162 IleLeuValLysThrGluAspGluAlaLysArgIleIleSerGluIleAspLysGlnPro 181
 239547 ATCTTAGTAGGACGACGAAAAAGAACTAGATATATTATTAACGAACTTAAGCTTTAAA 239606
 182 LysAlaLysLysGluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsn 201
 239607 CGTAAGAAGCTAGATGCTAAATTTAGCGAGCTTCTAAAGAGAAATCAATTGATCCAGCT 239666
 202 SerLysAsnAlaGlnAsnGlyLysAspLeuGlyLysPheGlnLysAsnGlnMetAlaPro 221
 239667 TCAAAAAC-----CAAGGTGAGTGGCTTGGTTGATCAATCAATGATGTAAG 239720
 222 AspPheSerLysAlaAlaPheAlaLeuThrProGluLysAspTyThrLysThrProValLys 241
 239721 CCTTTACAGATGCTGCTTTCGCGCTTAAAAATGATCATATTAACAATCCGGTTAAA 239780
 242 ThrGluPheGlyTyThrIleIleIleTyLeuIleSerLysAspSerProValThrTyThr 261
 239781 ACGAATTTTGGTTATCATGATATCTTAAAGAAATTCGCAAGCTAAAGTCAATCAAA 239840
 262 TyrGluGlnAlaLysProThrLysLysGluMetLeuGlnGluLysLeuPheGlnGluArg 281
 239841 TTGTAGTAAGTAATAACAGATATTTGAAAAACGACCTTAATTTGAAGAAATTTAAAAAGTT 239900
 282 MetAsnGlnArgIleGluGluLeuArgLysHisAlaLysIle 295
 239901 ATCAATCAAAAAGCGCAAGATCTCTTAATAGTGTAAAGT 239942

Search completed: April 24, 2005, 09:13:30
 Job time : 4774.41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2005, 02:20:20 ; Search time 542.741 Seconds
(without alignments)
3261.231 Million cell updates/sec

Title: US-10-039-183A-2
Perfect score: 1520
Sequence: 1 MKKNTLNALVGLSTSLM.....ERMGRIRERAKIVINK 299

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+g2n.model -DEV=xlh
-Q=/cgn2.1/USPTO/US10039183/runac_22042005_122424_20098/app_query.fasta_1.1038
-DB=N.GeneSeq_16Dec04 -QFMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORMext -HARSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10039183 -CCGN_1_1_796 @runac_22042005_122424_20098 -NCPU=6 -ICPU=3
-NO MMAP -LARGECQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.GeneSeq_16Dec04: *
1: GeneSeqn1980s: *
2: GeneSeqn1990s: *
3: GeneSeqn2000s: *
4: GeneSeqn2001as: *
5: GeneSeqn2002as: *
6: GeneSeqn2002as: *
7: GeneSeqn2002bs: *
8: GeneSeqn2003as: *
9: GeneSeqn2003bs: *
10: GeneSeqn2003cs: *
11: GeneSeqn2003ds: *
12: GeneSeqn2004as: *
13: GeneSeqn2004bs: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1520	100.0	1000	2	AAx14041 H. pylori
2	1520	100.0	1149	2	AAV07963 Helicobac
3	1520	100.0	1149	10	AAV07963 Helicobac
4	1486	97.8	900	2	AAV24859 H. pylori
5	1482	97.5	1082	2	AAV90591 Nucleotid

6	1482	97.5	1117	2	AAV90580 Nucleotid
7	1474	97.0	1082	2	AAV90844 Nucleotid
8	1311	86.2	844	6	ABX66063 Helicobac
9	1017	66.9	616	2	AAV90546 Nucleotid
10	679	44.7	456	2	AAV24638 H. pylori
11	664	43.7	595	6	ABX66722 Helicobac
12	233	15.3	960	2	AAQ22604 Gene enco
13	224	14.7	993	13	ABN91511 Staphyloc
14	224	14.7	993	13	ADSO2896 Staphyloc
15	221.5	14.6	1720	6	ABQ70541 Staphyloc
16	221.5	14.6	43980	6	ABQ67192 Staphyloc
17	221.5	14.6	110000	6	ABQ69245_23
18	220.5	14.5	110000	6	ABA03041_22
19	220.5	14.5	110000	6	ABA03041_23
20	211.5	13.9	2097	11	ABD15031 Pseudomon
21	211.5	13.9	2331	11	ABD15205 Pseudomon
22	211.5	13.9	63563	4	AAE28546 Genomic f
23	207.5	13.7	1134	9	ADB11571 Altolococ
24	207.5	13.7	1134	9	ADB11573 Altolococ
25	207.5	13.7	110000	9	ADB12064_14
26	205.5	13.5	960	8	ABT14941 Pathogen
27	205.5	13.5	960	8	ACF72655 Staphyloc
28	205.5	13.5	1110	12	ADL04132 DNA encod
29	200.5	13.2	1163	6	ABQ68437 Staphyloc
30	191.5	12.6	110000	6	ABQ69245_14
31	191.5	12.6	110000	6	ABQ67195_2
32	190.5	12.5	110000	10	ADF77343_17
33	187.5	12.3	7963	2	AAV74479 Staphyloc
34	187.5	12.3	110000	6	ABA03041_14
35	182.5	12.0	6157	6	ABQ70939 Staphyloc
36	180.5	11.9	1776	12	ADL70177
37	180.5	11.9	110000	2	AAT42063_10
38	180	11.8	72750	3	AAAB1468 N. mening
39	180	11.8	349980	3	AAE21544 Neisseria
40	179	11.8	191996	13	ADT05647 Haemophil
41	177	11.6	1776	12	ADL70175
42	175	11.5	1776	12	ADL70174
43	175	11.5	110000	3	AAAB1490_02
44	174	11.4	825	11	ABD05214 Pseudomon
45	174	11.4	855	11	ABD05038 Pseudomon

ALIGNMENTS

RESULT 1	AAx14041	standard; DNA; 1000 BP.
ID	AAx14041	
XX	AAx14041;	
AC	AAx14041;	
XX		
DT	31-MAR-1999	(first entry)
XX		
DE	H. pylori	GHPO 1360 gene.
XX		
KW	GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;	
XX	peptic ulcer disease; ss.	
XX		
OS	Helicobacter pylori.	
XX		
FN	Key	Location/Qualifiers
FT	CDS	51..950
FT		/tag= a
XX		
PN	W09843478-A1.	
XX		
PD	08-OCT-1998.	
XX		
PF	01-APR-1998;	98WO-US006371.
XX		
PR	01-APR-1997;	97US-00833457.
XX		
PR	24-JUN-1997;	97US-00881227.
XX		
PR	29-JUL-1997;	97US-00902615.
XX		

PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PA (HUMA-) HUMAN GENOME SCI INC.

PI Kleantinous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;

XX WPI; 1998-542293/46.

DR P-PSDB; AAW98322.

XX New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastrointestinal diseases.

XX Claim 1; Page 431-432; 2054dp; English.

XX This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these infections,
CC including acute, chronic, and atrophic gastritis, and peptic ulcer
CC diseases, e.g. gastric and duodenal ulcers. They can also be used for the
CC production of antibodies. The products can also be used for detection and
CC diagnosis

XX SQ Sequence 1000 BP, 372 A, 167 C, 207 G, 254 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	2,93e-140	Length:	1000
Score:	1520.00	Matches:	299
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-10-039-183a-2 (1-299) x AAX14041 (1-1000)

```
QY 1 MetLysLYsAsnIleleuAsnleuValleuValGlyAlaLeuSerThiSerPheLeuWet 20
DB 51 ATGAAATAATATCTTAATTAGCGTTAGCGGTGCGTTGAGCAGCGCTTTTGGANG 110
QY 21 AlaLYSPRoAlAHIsAsnAlaAsnAlaATHrHisAsnThryblyerThrThraPseR 40
DB 111 GCTAAGCGCGCTCATACCGCAATATACGCTAGCGCATTAACAGAAAAAGAGCTGATCT 170
QY 41 SerAlaGlyValleuAlaThrValAspGlyArgProIleThrySerAspPheAspMet 60
DB 171 TCAGCAGCGGTTCGTCAGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 230
QY 61 IleLYSGlnArgAsnProAsnPheAspPheAspLYsLeuLYSGlnLYSGlnLYSGln 80
DB 231 ATTAAGCAACGAATCTTAATTGATTGTTGACCAAGCTTAAGAGAAAAAGAGAGCC 290
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGlnAsnGlnAlaLYSthrGlnLYSLeu 100
DB 291 TTGATTGATCAAGCTATTCGACCGCCCTGTGAGAAAAAGAGAGAAAAAGAGAAATTG 350
QY 101 AspSerThrProGlnPheLYSAlaMetMetGlnAlaValLYSGlnAlaLeuValGln 120
DB 351 GACAGCAGCTCCGAGATTATTAACGATGATGAGAACCGGCTTAATAAACAGGCTTTAGTGA 410
QY 121 PheTTPAlaLYSLYSGlnAlaGlnGlnValLYSlySvalGlnIleProGlnLYSGlnWet 140
DB 411 TTTTGGGGCTAAAAAACAGGCTGAGAGAGAGAAAAAGATCCCAATCCAGAAAAAGATG 470
QY 141 GlnAspPheThyArgAsnAlaAsnLYSAspGlnLeuPheValLYSGlnAlaHisAlaArg 160
DB 471 CAAGATTTTTCACAGCTAACAGATCGCTTTTGTTCACAGCTTAAGAGAGAGAGAGAG 530
QY 161 HisIleLeuValLYSthrGlnAspGlnAlaLYSArgIleIleSerGlnIleAspLYSGln 180
DB 531 CATATTTTTGTAAAAACGAGATGAGGCTTAACGATTTATTTCTGAGATTACAAACAG 590
QY 181 ProLYSAlaLYSLYSGlnAlaLYSAspPheIleGlnLeuAlaAsnArgAspThryIleAspPro 200
```

```
DB 591 CCAAGGCTAAAAAGAGCTAAATTCATTGATTAGCCAAATCGGATACGATTGATCCT 650
QY 201 AsnSerLYSAsnAlaGlnAsnGlyLYSAspLeuGlyLYSPheGlnLYSAsnGlnMetAla 220
DB 651 AACAGCAAGAGAGCGCAAAATGCGGTATTTGGGAAATTCAAAAGAACCAATGGCT 710
QY 221 ProAspPheSerLYSAlaAlaPheAlaLeuThrProGlyAspTYrThrySerThryProVal 240
DB 711 CCGATTTTTCAAAGCCGCTTCCTTAACCTCGGGGATTAACATTAACCCCTGTT 770
QY 241 LysThrGlnPheGlyTYrHisIleIleLYSLeuIleSerLYSAspSerProValThryr 260
DB 771 AAACAGAGTTGGTTATCATATTAATTTGATTGATTCTTAAGATAGCCCTGTATCTAT 830
QY 261 ThryrGlnGlnAlaLYSProThryrIleLYSGlyMetLeuGlnGlnLYSLeuPheGlnGln 280
DB 831 ACTTATGAACAGGCTTAACCTTAACATTAAGGGAGATTTACAGAAAGCTTTTTCAGAGA 890
QY 281 ArgMetAsnGlnArgIleGlnGlnLeuArgLYSAlaLYSAlaLYSAlaLYSAlaLYS 299
DB 891 CGCATGAATCAACGATTTGAGGAACTTAAGAAAGCAGCTTAATAATTGTTATCAACAG 947

RESULT 2
AAV07963
ID AAV07963 standard; DNA; 1149 BP.
XX
XX AC AAV07963;
XX
DT 25-MAR-2003 (revised)
DT 02-FEB-1999 (first entry)
DE Helicobacter pylori 32 kDa polypeptide GHPO 1360 DNA.
XX
XX GHPO 1360; infection; gastritis; ulcer; vaccine; diagnosis; therapy; ss.
XX
XX Helicobacter pylori.
XX
OS Helicobacter pylori.
XX
FH CDS
FT 106..1005
FT sig_peptide
FT /*tag= a
FT /*tag= b
FT mat_peptide
FT /*tag= c
XX
XX PN
XX MO9843479-A1.
XX
XX PD 08-OCT-1998.
XX
XX PF 31-MAR-1998; 98WO-US006421.
XX
XX PR 01-APR-1997; 97US-00831310.
XX
XX PR 01-APR-1997; 97US-00834666.
XX
XX PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Kleantinous H, Lissolo L, Tomb J, Miller C, Algarawi A;
XX
XX WPI; 1998-568251/48.
XX
XX DR P-PSDB; AAW73034.
XX
XX PT New isolated Helicobacter polynucleotides - used to develop products for
XX the diagnosis, prevention and treatment of Helicobacter infections and
XX PT gastroduodenal diseases.
XX
XX PS Claim 1; Page 148-149; 184dp; English.
XX
XX This DNA sequence codes for a 32 kDa Helicobacter pylori polypeptide (see
XX AAW73034) designated GHPO 1360. A polynucleotide encoding the
XX unprocessed, or the mature, GHPO 1360 can be obtained from genomic DNA by
XX PCR amplification (see AAV07963-70 and AAV07973). The invention provides
XX polynucleotides (see AAV72001, AAV07912-21 and AAV07963-64) encoding a
```

CC family of 76 kDa Helicobacter polypeptides (see AAW73022-32), GHPD 1360
CC and a 50 kDa polypeptide (see AAW73035). These polynucleotides were
CC initially identified in a search of H. pylori genomic databases. DNA
CC cases for expression of the Helicobacter proteins (unprocessed or
CC mature forms) in prokaryotic or eukaryotic cells are provided. The
CC polynucleotides can be used in vaccines to prevent or treat Hb infection
CC in a mammal. Viral (especially poxvirus) or bacterial vectors are used.
CC Products and methods of the invention allow treatment and prevention of
CC gastroduodenal diseases associated with Hb infections, including acute,
CC chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric
CC and duodenal ulcers. Diagnostic and detection methods are also provided.
CC GHPD 1360 was demonstrated to be a protective antigen. (Updated on 25-MAR
CC -2003 to correct PI field.)

XX Sequence 1149 BP; 412 A; 185 C; 242 G; 310 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3,45e-140	Length:	1149
Score:	1520.00	Matches:	299
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-10-039-183A-2 (1-299) x AAV07963 (1-1149)

QY 1 MetLysLYAsnLIleuSbnLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 106 ATGAAATAAAATATCTTAAATTAGCGTTAGCGGCGCTTAAAGCAAGCTGCTTTTGAAG 165
QY 21 AlAlYSPrcAlAHlSbnAlAsbnAlaThRHlSbnThRlySerThrRhpSer 40
DB 166 GCTAAGCCGGCTCATACCAATAACGTAACGCAATACCAAGAAAAAGCACTGATTCT 225
QY 41 SerAlGlyValLeuAlaThRValAspGlyAProIleThrLysSerAspPheAspMet 60
DB 226 TCAGCAGCGGTGTGCGACAGTGTGCGACAGCTATCATTAAGCAATTTTGACATG 285
QY 61 IleLYSGlnARsbnProbnAspPheAspPheAspLysLeuLYSGlnLYSGlnLY 80
DB 286 ATTAAGCAAGCAAAATCTTAATTTGTTGACAGCTTAAAGCAAGAAAAAGAAAGCC 345
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGlyAsnGlnAlaIleThrGlyLysLeu 100
DB 346 TTGATTGATCAAGCTATTTCGCAACGCCCTTGTAGAAAAAGAAAGCAAGAAATTG 405
QY 101 AppSerThrProGluPheLYSAlaMetMetGlnAlaValLYSGlnAlaLeuValGlu 120
DB 406 GACAGCACTCCAGAAATTTAAAGCATGATGAAAGCGGTTAAAGCAAGCTTTAGTGGAA 465
QY 121 PheTTPAlaLYSAlaLYSGlnAlaGluGluValLYSAlaGlnIleProGluLYSGlnMet 140
DB 466 TTTTGGGCTTAAAGAAACAGGCTGAAAGAAAGAAAGTCCAATCCCAAGAAAGAAAG 525
QY 141 GlnAspPheTYAsnAlaAsnLYSAspGlnLeuPheValLYSGlnGlnAlaIleAlaArg 160
DB 526 CAAAGATTTTCAACGCTAACAAAGATCAGCTTTTGTCAAGCAAGAACCCATGCTAGG 585
QY 161 HisIleLeuValLYSAlaLYSGlnAlaLYSAlaLYSAlaLYSAlaLYSAlaLYSAla 180
DB 586 CATATTTTGTGAAACCGAAGATGAGGCTTAAAGCAATTTTGTGATTAACAAACAG 645
QY 181 ProLYSAlaLYSAlaLYSGlnAlaLYSAlaLYSAlaLYSAlaLYSAlaLYSAla 200
DB 646 CCAAAAGGCTTAAAGAAAGCAATTTCAATGAGTTAGCCAAATCGGATTCGATTGATCT 705
QY 201 AsnSerLYSAsnAlaGlnAsnGlyLYSAspLeuGlyLYSAspPheGlnLYSAsnGlnMetAla 220
DB 706 AACACCAAGAAACCGCAAAATGCGGTGATTTGGGGAATTTCCAAAGAACCAATAGCT 765
QY 221 ProAspPheSerLYSAlaAlaPheAlaLeuThrProGlyAspTYRThrLYSAlaProVal 240
DB 766 CCGGATTTTCTTAAAGCCGCTTCTGCTTAACCTCGGGGATTAACCTAAACCCCTGTT 825

QY 241 LysThrGluPheGlyTYRHisIleIleIleTYRLeuIleSerLYSAspSerProValThrTYR 260
DB 826 AAAACAGATTTGCTTATCATATTTATTTGATTTTCAAGATAGCCCTGATTAAT 885
QY 261 ThrTYRGlulAlaLYSProThrIleLYSGlyMetLeuGlnGluLYSLeuPheGlnGlu 280
DB 886 ACTTATGAACAGGCTTAAACCTTCAATTAAGGGAATTTTCAAGAAAGCTTTCCAAAGA 945
QY 281 ArgMetAsnGlnARsbnIleGluGluLeuArgLYSAlaLYSAlaLYSAlaLYSAlaLYS 299
DB 946 CGCATGAATCAACGATTAAGAAAGCAAGCTTAAATTTGTTATCAACAAG 1002

RESULT 3

AAD61576 standard, DNA; 1149 BP.

AC AAD61576;

DT 15-JUN-2004 (first entry)

DE Helicobacter pylori p32 antigen DNA.

XX Immune response; Helicobacter infection; adjuvant; heat-labile toxin; LT;

KW cholera toxin; CT; urease; therapy; antibacterial; vaccine; antigen; p32;

KW gene; ds.

OS Helicobacter pylori.

XX Key Location/Qualifiers

FT CDS 106..1005

FT /tag= a

FT /product= "Antigen p32"

FT sig_peptide 106..166

FT /tag= b

FT mat_peptide 167..1002

FT /tag= c

FT /product= "Mature p32"

XX US6576244-B1.

XX 10-JUN-2003.

XX 18-JUN-1999; 99US-0036115.

XX 19-JUN-1998; 98US-00100258.

XX (ACAM-) ACAMBIS INC.

XX Wetzlin RA, Guy B;

XX WPI; 2003-799824/75.

XX P-PSDB; ABW00787.

XX Inducing immune response to Helicobacter useful for treating Helicobacter

PT pylori infection, by administering immunogenic Helicobacter polypeptide

PT admixed with adjuvant having heat-labile toxin of Escherichia coli.

XX Claim 8; Col 105-108; Opp; English.

XX The invention relates to a method for inducing protective immune response

CC to Helicobacter infection. The method comprising administering to a

CC mammal by injection an immunogenic Helicobacter pylori polypeptide

CC comprising a subunit of H. pylori urease admixed with an adjuvant having

CC one or more heat-labile toxin of Escherichia coli (LT), B subunit of LT

CC (UTB), cholera toxin (CT), and B subunit of CT. The method is useful for

CC inducing an immune response to Helicobacter infection in a mammal. The

CC method is useful for both treatment and prevention of H. pylori infection.

CC The invention is useful as vaccine. The present sequence is Helicobacter

CC pylori p32 antigen DNA

XX Sequence 1149 BP; 412 A; 185 C; 242 G; 310 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,45e-140 Length: 1149
 Score: 1520.00 Matches: 299
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-039-183a-2 (1-299) x AAD61576 (1-1149)

```

QY 1 MetLysLysAsnLLeuLLeuLLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 106 ATGAAAAAAATATCTTAATATTAGCGGTAGGGGCGCTTAGCGACGCTTTTGGATG 165
QY 21 AlAlAsProAlAhIsAsnAlAsnAsnAlaThrHiAsnThrLysThrThrAspSer 40
DB 166 GCTAAGCCGGCTCATACCCAAATTAACGCTACGATTAACGAAAAAAGCATGATTCT 225
QY 41 SerAlaGlyValLeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMet 60
DB 226 TCAGCAGGCGGTGTAGCGACAGTGGATGGCAGACCTATCACTAAAAAGCATTTTGCATG 285
QY 61 IleLysGlnArgAsnProAsnPheAspPheAspLysLeuLysGlnLysGlnLysGlnAla 80
DB 286 ATTTAAGCAACGAATCTTAATTTTGAATTTTGAACAACCTTAAAGAGAAAAAGAAAGAC 345
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGlnLysLeu 100
DB 346 TTGATTTGATCAACGCTATTGCGACCGCGCTTGTGAAAAAGAAAGCTTAAACCGAAGAA 405
QY 101 AspSerThrProGlnLysLysAlaMetMetGlnAlaValLysGlnAlaLeuValGlu 120
DB 406 GACACACATCCAGAAATTTAAAGCATGATGAGAAAGCGCTTAAAGACGCTTTAGTGAA 465
QY 121 PheThrAlaLysLysGlnAlaGlnGluValLysLysValGlnIleProGlnLysGlnMet 140
DB 466 TTTTGCGCTTAAAGAAAGCGCTGAGAAAGTGAAGAAAAAGTCCAAATCCCAAAAAAGAAAG 525
QY 141 GlnAspPheThrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArg 160
DB 526 CAAGATTTTTCACCGCTTACCAAGATCAGCTTTTGTGACGACGAAAGCCATGCTAGG 585
QY 161 HisIleLeuValLysThrGlnLysGlnAlaLysArgIleIleSerGlnIleAspLysGln 180
DB 586 CATATTTTATGTAAGAAACCGAAGATGAGCTTAAACGATATTTCTGAGATTGACAAACAG 645
QY 181 ProLysAlaLysLysGlnAlaLysPheIleGlnLeuAlaAsnArgAspThrIleAspPro 200
DB 646 CCAAAAGGCTTAAAGAAAGAAAGCTTAAATTCATGAGTTAGCCAAATCGGGATTCGATTCCT 705
QY 201 AsnSerLysAsnAlaGlnAsnGlyLysAspLeuGlyLysPheGlnLysAsnGlnMetAla 220
DB 706 AACACGAAACGCGCCAAATATGCGGTGATTTGGGGAATTTCCAAAGAACCAATATGGCT 765
QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspLysThrLysThrProVal 240
DB 766 CCGGATTTTCTTAAAGCCGCTTGGCTTAACTCCGCGGATTAACCTTAAACCCCTGTT 825
QY 241 LysThrGlnPheGlyThrHisIleIleLysLysLysLysLysLysLysLysLysLys 260
DB 826 AAAACAGAGTTGGTATCATATTTATTTATTTGATTTTCTTAAAGATTAAGCCCTTAACTAT 885
QY 261 ThrTYrGlnGlnAlaLysProThrIleLysGlnMetLeuGlnGlnLysLeuPheGlnGlu 280
DB 886 ACTTATGAAACAGGCTTAAACCTTACCTTAAAGGGAATGTTTCAAGAAAGCTTTTCCAAAGA 945
QY 281 ArgMetAsnGlnArgIleGlnLysLeuArgLysHisAlaLysIleValIleAsnLys 299
DB 946 CGCATGATCAACGACATTAAGAAAGCTTAAAGAAAGCAACGCTTAAATTTGTTATCAACAAG 1002

```

RESULT 4
 AAV24859
 ID AAV24859 standard; DNA; 900 BP.

```

XX AC AAV24859;
XX XX
XX 24-JUN-1998 (first entry)
XX DT
XX DT
XX H. pylori cell envelope OMP ORF 02ae11612_22477267_f2_27.
XX XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bacteria; life cycle; activator;
XX inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
XX Heliobacter pylori.
XX OS
XX Key Location/Qualifiers
XX FH 1..900
XX FT /*tag= a
XX FT
XX MO9737044-A1.
XX PD
XX 09-OCT-1997.
XX XX
XX 27-MAR-1997; 97MO-US005223.
XX PF
XX 29-MAR-1996; 96US-00625811.
XX PR 02-APR-1996; 96US-00758731.
XX PR 25-OCT-1996; 96US-00736905.
XX PR 28-OCT-1996; 96US-00738859.
XX PR 06-DEC-1996; 96US-00761318.
XX PA (ASTR ) ASTRA AB.
XX PI Smith D, Alm RA;
XX XX
XX WPI: 1997-503122/46.
XX DR P-PSDB; AAM55450.
XX XX
XX Helicobacter pylori nucleic acid sequences and encoded polypeptide(s) -
XX PT useful in vaccines to treat or prevent H. pylori infection and for
XX PT diagnosis of H. pylori infection.
XX XX
XX Claim 5,6,21; Page 288; 1145pp; English.
XX XX
XX This sequence encodes a H. pylori cell envelope outer membrane protein
XX (OMP) having no terminal Phe residue. The protein may be used in a
XX vaccine to prevent or treat H. pylori infection or to identify H. pylori
XX polypeptide binding compounds, useful as potential H. pylori life cycle
XX activators or inhibitors. The DNA and probes derived from it may be used
XX for the identification of H. pylori in a sample and the diagnosis of H.
XX CC pylori infection. Nucleic acid sequences complementary to the DNA act as
XX CC antisense sequences and can be used to prevent the translation of H.
XX CC pylori mRNA. Antibodies against the protein can be used in immunoassays
XX CC to evaluate the abundance and distribution of H. pylori-specific
XX CC antigens. The genomic sequence of H. pylori (ATCC 55679) was determined
XX CC from overlapping contigs generated by mechanically shearing the bacterial
XX CC DNA. The sequences were analysed for ORF of at least 180 nucleotides, and
XX CC the predicted coding regions defined by computer evaluation. To identify
XX CC likely H. pylori antigens for vaccine development, the amino acid
XX CC sequences predicted from various ORF were analysed for significant
XX CC homology to other known or exported membrane proteins. Having identified
XX CC and determined the sequences of interest, particular regions can be
XX CC isolated from H. pylori by PCR amplification for recombinant polypeptide
XX CC production, e.g. in E. coli hosts
XX XX
XX Sequence 900 BP; 329 A; 161 C; 193 G; 217 T; 0 U; 0 Other;

```

Alignment Scores:

Pred. No.: 5.77e-137 Length: 900
 Score: 1486.00 Matches: 290
 Percent Similarity: 99.00% Conservative: 6
 Best Local Similarity: 96.98% Mismatches: 3
 Query Match: 97.76% Indels: 0
 DB: 2 Gaps: 0

US-10-039-183A-2 (1-299) x AAV24859 (1-900)

```
QY 1 MetlyslvsaenllleuasnleuallaleuvalglvalaleuSerThSerPheleuMet 20
DB 1 ATGAAAAAAATATCTTAATTTACGGTTAGTGGCGGCTGAGAGCGCGCTTTTGTGATG 60
QY 21 AlAlvSPProAlAHleAsnAlaAsnAsnAlaThRHisAnThRlyLeThThraAspSer 40
DB 61 GCTAAGCCGGCTCATATCCGAAATACCTACGCTAACACCAAGAAAGAAACGCTATGCT 120
QY 41 SerAlaGlValleuAlaThRValAspGlYArGProlleThRlySerAspPheAspMet 60
DB 121 TCAGCAGCGGTGTAGCAGAGTGATGAGCAGACCCATCACCAAGAAAGCATTTGATATG 180
QY 61 lIelySGlnArGAsnProAsnPheAspPheAspPheAspPheAspPheAspPheAspPhe 80
DB 181 ATTAAGCAACGAAATCTTAATTTGATTTGCAAGCTTAAAGAAAGAAAGAAAGAAAGCC 240
QY 81 LeuIlleAspGlnAlaAlaIleArGThRAlaLeuValGluaAsnGluaAlaIleThRGlulysleu 100
DB 241 TTGATTGACAAAGCTATCCGACCGCGCTTGAGAAATAGAGCTTAAGCAGAAAGCTT 300
QY 101 AspSerThRProGlupHelysAlaMetMetGluaAlaValIlylSGlnAlaLeuValGlu 120
DB 301 AATCAGACTCCAGAAATTTAAAGCGATGATGAGAAAGCGTTAAAGAAAGCGCTTATGAGAA 360
QY 121 PheTTPAlaIlylSGlnAlaGlulGluValIlylSGlnValGlnIleProGluIlySGlnMet 140
DB 361 TTTTGGGCTTAAAGAAAGCGCTGAGAGAGTGAAGAAAGTCCAAATCCAGAAAGAAAGAAAGT 420
QY 141 GlnAspPheThRAsnAlaAsnlyAspGlnleuPheAlIlySGlnAlaAlaHisAlaArg 160
DB 421 CAGGATTTTACACGCGCAATTAACATCAAGCTTTTTCACACAGAAAGCCATCTAGG 480
QY 161 HisIleleuValIlyThRGluaSpGlnAlaIyAsArgIleIleSerGlnIleAspGln 180
DB 481 CATATTATTAGTGAAGAAAGCGAAGATGAGACCAAGAAAGATTTATTTGAGATTGACAAACAG 540
QY 181 ProIySAlaIlylSGlnAlaIlyPheIleIleGluLeuAlaAsnArGAspThRlIleAspPro 200
DB 541 CCAAGGCTTAAAGAAAGAAAGCCAAATTCATGATTAGCCAAATCGGAGATACGATTCCT 600
QY 201 AsnSerIySAsnAlaGluaSngIyGlIyAspLeuGlyIySPheGlnIyAsnGlnIleAla 220
DB 601 AACGCAAGAAAGCGCCAAATGCGCGTATTTGGGAAATTTCCAAAGAAAGCAAAATGGCT 660
QY 221 ProAspPheSerIySAlaAlaPheAlaLeuThRProGlyAspThRThlySerThRProVal 240
DB 661 CCGGATTTTCTTAAAGCGCGCTTGCCTTTAACTCTCGGGAGTTACCTTAAAGCCCTGTT 720
QY 241 LysThRGlupHelyThRHisIleIleIyRleuIleSerIyAspSerProValThRlyR 260
DB 721 AAAACAGGTTTGGTATCATATTTATTTGATTTTAAAGATAGCGCTGTAATCTTAT 780
QY 261 ThRlyRGlupHelyAlaIyProThRlIleIySGlyMetleuGlnIlyLeuPheGlnIly 280
DB 781 ACTTATGCAAGCTTAACCTTACCTTAAAGGGAGATTTTCAAGAAAGAAAGCTTTTCCAGAA 840
QY 281 ArgMetAsnGlnArGileGluGluLeuArGlySHisAlaIyIleValIleAsnlyS 299
DB 841 CGCATGATCAACGCAATTTAGAGAAATTTAAGAAAGCAAGCTTAAATTTGTTATCAACAG 897
```

RESULT 5
AAV90591 ID AAV90591 standard; DNA; 1082 BP.XX AAV90591;
XX AC
XX 20-MAR-2003 (revised)
XX 18-FEB-1999 (first entry)
XX Nucleotide sequence of clone Y104.1.ASM from cluster 3.
DE
XX

```
KM Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;  
KM peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.  
XX Helicobacter pylori.  
XX OS  
XX W09849314-A2.  
XX PN  
XX 05-NOV-1998.  
XX PD  
XX 25-APR-1998; 98WO-US008487.  
XX PF  
XX 25-APR-1997; 97US-0045107P.  
XX PR  
XX 14-OCT-1997; 97US-0061958P.  
XX PR  
XX (GENE-) GENELABS TECHNOLOGIES INC.  
XX PA  
XX Chow TP, Fry KE, Lim MY, Mcatee CP,  
XX WPI, 1999-009433/01.  
XX DR  
XX New Helicobacter pylori antigens and related nucleic acid sequences -  
XX PT useful in serological diagnosis and protective vaccines, providing long-  
XX PT lasting immune response.  
XX PS  
XX Claim 27; Page 123-124; 402pp; English.  
XX  
XX The present sequence encodes a Helicobacter pylori antigenic protein that  
XX CC is characterized by immunoreactivity with H. pylori-positive antisera.  
XX CC The proteins are highly immunogenic and induce a long-lasting immune  
XX CC response that persists even after antimicrobial treatment. In antibody-  
XX CC detection assays, on sera, plasma, urine, saliva etc., they are highly  
XX CC sensitive and specific. The specification also describes 69 previously  
XX CC unrecognized immunogenic cluster families. H. pylori antigens are used to  
XX CC detect H. pylori-specific antibodies, for diagnosing infection or to  
XX CC confirm eradication of infection, and in vaccines to protect against H.  
XX CC pylori infection and related diseases (gastritis, peptic ulcer, gastric  
XX CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)  
XX  
SQ Sequence 1082 BP; 383 A; 189 C; 222 G; 268 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 1-786-136 Length: 1082  
Score: 1482.00 Matches: 291  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.32% Mismatches: 5  
Query Match: 97.50% Indels: 0  
DB: 2 Gaps: 0  
  
US-10-039-183A-2 (1-299) x AAV90591 (1-1082)  
QY 1 MetlyslvsaenllleuasnleuallaleuvalglvalaleuSerThSerPheleuMet 20  
DB 1 ATGAAAAAAATATCTTAATTTACGGTTAGTGGCGGCTGAGAGCGCGCTTTTGTGATG 224  
QY 21 AlAlvSPProAlAHleAsnAlaAsnAsnAlaThRHisAnThRlyLeThThraAspSer 40  
DB 225 GCTAAGCCGGCTCATATCCGAAATACCTACGCTAACACCAAGAAAGAAACGCTATGCT 284  
QY 41 SerAlaGlValleuAlaThRValAspGlYArGProlleThRlySerAspPheAspMet 60  
DB 285 TCAGCAGCGGTGTAGCAGAGTGATGAGCAGACCCATCACCAAGAAAGCATTTGATATG 344  
QY 61 lIelySGlnArGAsnProAsnPheAspPheAspPheAspPheAspPheAspPheAspPhe 80  
DB 345 ATTAAGCAACGAAATCTTAATTTGATTTGCAAGCTTAAAGAAAGAAAGAAAGAAAGCC 404  
QY 81 LeuIlleAspGlnAlaAlaIleArGThRAlaLeuValGluaAsnGluaAlaIleThRGlulysleu 100  
DB 405 TTGATTGACAAAGCTATCCGACCGCGCTTGAGAAATAGAGCTTAAGCAGAAAGCTT 464  
QY 101 AspSerThRProGlupHelysAlaMetMetGluaAlaValIlylSGlnAlaLeuValGlu 120  
DB 465 GATCAGACTCCAGAAATTTAAAGCGATGATGAGAAAGCGTTAAAGAAAGCGCTTATGAGAA 524
```

```
QY 121 PheTPAlaLysLysGlnAlaGluValLysLysValGlnIleProGluLysGluMet 140
DB 525 TTTTGGGCTAAAAAAGAGCTGTAAGAGTGAAGAAAAAGTCCAAATCCCAAAAAAGAAAG 584
QY 141 GlnAspPheTYrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArg 160
DB 585 CAAGATTTTTCACAGCTTAATTAAGATCAGCTTTTGTTCACACAGAACCCCATGCTAGG 644
QY 161 HisIleLeuValLysThrGluAspGlnAlaLysArgIleIleSerGlnIleAspLysGln 180
DB 645 CATATTTTAGTGAAGAACGGAAGATGAGCTAAACGATTAATTTCTGAGATTGACAAACAG 704
QY 181 ProLysAlaLysLysGlnAlaLysPheIleGlnLeuAlaAsnArgAspThrIleAspPro 200
DB 705 CCAAAAGGCTAAAAAGAACCCAAATTCAATTGAGTTAGCCAAATCGGATACGATTGATCCT 764
QY 201 AsnSerLysAsnAlaGlnAsnGlyGlyAspLeuGlyLysPheGlnLysAsnGlnMetAla 220
DB 765 AACACGACAGAACGCGCAAAATGCGGTGATTTGGGGAAATTCCAAAAGAACCAAAATGCGCT 824
QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTYrThrLysThrProVal 240
DB 825 CCGGATTTTCTAAAGCGCTTTCGCTTAACCTCTGGGGATTTACCTAAACCCCTGTT 884
QY 241 LysThrGluPheGlyTYrHisIleIleTYrLeuIleSerLysAspSerProValThrTYr 260
DB 885 AAAACAGATTTGGTTATCATATTAATCTTTGATTTCTTAAAGATAGCCCTGTAATCTTAT 944
QY 261 ThrTYrGluGlnAlaLysProThrIleLysGlyMetLeuGlnLysLeuPheGlnGlu 280
DB 945 ACTTATGAGCAAGCTTAACCTCACTTAAGGGAGATGTTCAAGAAAGCTTTTCCAAAGAA 1004
QY 281 ArgMetAsnGlnArgIleGlnLysLeuArgLysHisAlaLysIleValIleAsnLys 299
DB 1005 CGCATGAATCAACGACTGAGAAATTAAGGAAGACAGCTAAATGTTCTCAACAG 1061

RESULT 6
AAV90580
ID AAV90580 standard; DNA; 1117 BP.
XX
AC AAV90580;
XX
DT 20-MAR-2003 (revised)
DT 18-FEB-1999 (first entry)
XX
DE Nucleotide sequence from clone Y104-1.asm.
XX
KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX
OS Helicobacter pylori.
XX
FN WO9649314-A2.
XX
PD 05-NOV-1998.
XX
PF 25-APR-1998; 98WO-US008487.
XX
PR 25-APR-1997; 97US-0045107P.
PR 14-OCT-1997; 97US-0061958P.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Chow TP, Fry KE, Lim MY, Mcatee CP;
XX
DR WPI, 1999-009433/01.
DR P-PSDB; AAW89829.
XX
PT New Helicobacter pylori antigens and related nucleic acid sequences -
PT useful in serological diagnosis and protective vaccines, providing long-
PT lasting immune response.
XX
```

```
PS Claim 27; Page 119; 402pp; English.
XX
CC The present sequence encodes a Helicobacter pylori antigenic protein that
CC is characterised by immunoreactivity with H. pylori-positive antisera.
CC The proteins are highly immunogenic and induce a long-lasting immune
CC response that persists even after antimicrobial treatment. In antibody-
CC detection assays, on sera, plasma, urine, saliva etc., they are highly
CC sensitive and specific. The specification also describes 69 previously
CC unrecognized immunogenic cluster families. H. pylori antigens are used to
CC detect H. pylori-specific antibodies, for diagnosing infection or to
CC confirm eradication of infection, and in vaccines to protect against H.
CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
CC
SQ Sequence 1117 BP; 388 A; 202 C; 234 G; 293 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,85e-136 Length: 1117
Score: 1482.00 Matches: 291
Percent Similarity: 98.33% Conservative: 3
Best Local Similarity: 97.32% Mismatches: 5
Query Match: 97.50% Indels: 0
DB: 2 Gaps: 0

US-10-039-183a-2 (1-299) x AAV90580 (1-1117)
QY 1 MetLysLysAsnIleLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 183 ATGAAAAAATAATTTTAAATTTAGCTTAGCGGCGGCTTACGTCGCTTTTGATG 242
QY 21 AlaLysProAlaHisAsnAlaAsnAlaThrHisAsnThrLysLysThrThraAspSer 40
DB 243 GCTAAGCCGCGCTCAATACGACAGATACCTACGCTAACACCAAAAAAAGCATATTCT 302
QY 41 SerAlaGlyValIleAlaIleThrValAspGlyArgProIleThrLysSerAspPheAspMet 60
DB 303 TCACCCGCGGTGTCGACAGTGGATGACGACCTTACCTAAAGCGCATTTTATATG 362
QY 61 IleLysGlnArgAsnProAsnPheAspPheAspLysLeuLysGlnLysGlnAla 80
DB 363 ATTAAAGCAAGAAATCCTAATTTTGAATTTTGAACAAGCTTAAAGAAAGAAAAAGAAAGCC 422
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGlnAsnGlnAlaLysThrGluLysLeu 100
DB 423 TTGATTTGACAAAGCTATCCGACCGCACTTGAAGAAATGAGGCTAAGGCAAGAAAAAGCTC 482
QY 101 AspSerThrProGluPheLysAlaMetMetGlnAlaValLysLysGlnAlaLeuValGlu 120
DB 483 GATCAGACTCCAGATTTTAAAGCGATGATGAAAGCGTTAAAAACAGGCTTTAGTGGAA 542
QY 121 PheTPAlaLysLysGlnAlaGluValLysLysValGlnIleProGluLysGluMet 140
DB 543 TTTTGGGCTTAAAAAAGAGCTGTAAGATGAAAAAGTCCAAATCCGAAABAAAAGATG 602
QY 141 GlnAspPheTYrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArg 160
DB 603 CAAGATTTTTCACAGCTTAATTAAGATCAGCTTTTGTTCACAGCAAGAGCCATCTAGG 662
QY 161 HisIleLeuValLysThrGluAspGlnAlaLysArgIleIleSerGlnIleAspLysGln 180
DB 663 CATATTTTAGTGAAGAACGGAAGATGAGCTAAACGATTAATTTCTGAGATTGACAAACAG 722
QY 181 ProLysAlaLysLysGlnAlaLysPheIleGlnLeuAlaAsnArgAspThrIleAspPro 200
DB 723 CCAAAAGGCTAAAAAGAACCCAAATTCAATTGAGTTAGCCAAATCGGATACGATTGATCCT 782
QY 201 AsnSerLysAsnAlaGlnAsnGlyGlyAspLeuGlyLysPheGlnLysAsnGlnMetAla 220
DB 783 AACACGACAGAACGCGCAAAATGCGGTGATTTGGGGAAATTCCAAAAGAACCAAAATGCGCT 842
QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTYrThrLysThrProVal 240
DB 843 CCGGATTTTCTAAAGCGCTTTCGCTTAACCTCTGGGGATTTACACTAAACCCCTGTT 902
```


QY 241 LysThrGluPheGlyTyrHisIleIleTyrLeuIleSerIysAspSerProValThrTyr 260
DB 903 AAAACAGAGTTGGTTATCATATTTATCTATTTGATTTCTAAAGATAGCCCTGTAACCTTAT 962
QY 261 ThrTyrGluGlnAlaIleProThrIleIleGlyMetLeuGlnGluIlePheGlnGlu 280
DB 963 ACTTATGAGCAAGCTTAACCTACATTAAGGGGATGTTACAGAAAGCTTTCCAGAGA 1022
QY 281 ArgMetAsnGlnArgIleGluGluLeuArgIleHisAlaIleValIleAsnIys 299
DB 1023 CGCATGAATCAACGATTAAGAGATTAAAGAAAGACGCTAAATGTTCTCAACAG 1079
RESULT 7
AAV90844 standard; DNA; 1082 BP.
ID AAV90844
XX AAV90844;
AC AAV90844;
XX 20-MAR-2003 (revised)
DT 18-FEB-1999 (first entry)
XX Nucleotide sequence of cluster 3.
DE Nucleotide sequence of cluster 3.
XX Antigen: immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
OS Helicobacter pylori.
XX MO9849314-A2.
XX 05-NOV-1998.
XX 25-APR-1998; 98MO-US008487.
XX 25-APR-1997; 97US-0045107P.
PR 14-OCT-1997; 97US-0061958P.
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX Chow TP, Fry KE, Lam MY, Mcatee CP;
PI WPI, 1999-009433/01.
XX New Helicobacter pylori antigens and related nucleic acid sequences -
PT useful in serological diagnosis and protective vaccines, providing long-
XX lasting immune response.
PS Claim 20; Page 281-282; 402pp; English.
XX The present sequence encodes a Helicobacter pylori antigenic protein that
CC is characterised by immunoreactivity with H. pylori-positive antisera.
CC The proteins are highly immunogenic and induce a long-lasting immune
CC response that persists even after antimicrobial treatment. In antibody-
CC detection assays, on sera, plasma, urine, saliva etc., they are highly
CC sensitive and specific. The specification also describes 69 previously
CC unrecognised immunogenic cluster families. H. pylori antigens are used to
CC detect H. pylori-specific antibodies, for diagnosing infection or to
CC confirm eradication of infection, and in vaccines to protect against H.
CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
XX Sequence 1082 BP; 383 A; 188 C; 222 G; 286 T; 0 U; 3 Other;
Alignment Scores:
Pred. No.: 1.09e-135 Length: 1082
Score: 1474.00 Matches: 290
Percent Similarity: 97.66% Conservative: 2
Best Local Similarity: 96.99% Mismatches: 7
Query Match: 96.97% Indels: 0
DB: 2 Gaps: 0
US-10-039-183A-2 (1-229) X AAV90844 (1-1082)

QY 1 MetIlyAsnIleLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 165 ATGAAAAAATATTTTAAATTTAGCGGGCGCGGTAGCGCGCTTTTGGATG 224
QY 21 AlaIysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrThrAspSer 40
DB 225 GCTAAGCCGGCTCATTAAGCAGATAGTACCTAGCTAATACCAAAAAAGCATGTTCT 284
QY 41 SerAlaGlyValLeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMet 60
DB 285 TCACCCGGCGGTGGTCAGACAGTGTAGCAGACACTTACCTAAAGCGATTTGATAG 344
QY 61 IleIysGlnArgAsnProAsnPheAspPheAspLysLysGluIleGlyGluAla 80
DB 345 ATTAAGCAACGAATCTTAATTTTATTTTGAACAGCTTAAGAGAAAGAAAGAAAGCC 404
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGluAlaIleThrGluIleLeu 100
DB 405 TTGATTGAGCAAGCTATCGGACCGCAGCTTGTAGAAAATGAGGCTTAAGCAGAAAGCTC 464
QY 101 AspSerThrProGluPheLysAlaMetMetGluAlaValIleLysGlnAlaLeuValGlu 120
DB 465 GATCAGACTCCAGAAATTTAAAGCGATGATGAGAGCGGTTAAAGAAAGCGCTTTAGTGGA 524
QY 121 PheTPAlaLysLysGlnAlaGluGluValIleLysLysValGlnIleProGluIleGluMet 140
DB 525 TTTGGGCTAAAGAAAGAGCTGAGAGAGTGAAGAAAGTCCAAATCCAGAAAGAAAGATG 584
QY 141 GlnAspPheTyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArg 160
DB 585 CAGATTTTTCACAGCTTAATTAAGATCAGCTTTTGTGCAAGCAAGAGCCATGCTAGG 644
QY 161 HisIleLeuValLysThrGluAspGluAlaLysValGlleIleSerGluIleAspLysGln 180
DB 645 CATATTATTAGTGAAGAAAGCAAGATGAGCTAAAGCAATTAATTTGAGATTGACAAACAG 704
QY 181 ProLysAlaLysLysGluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspPro 200
DB 705 CCAGAGGCTAAAGAAAGAGCCAAATTCATGATGATGCAATCGGATACGATTGATTCCT 764
QY 201 AsnSerLysAsnAlaGlnAsnGlyIleLysPheGlyLysPheGlnLysAsnGlnMetAla 220
DB 765 AACAGCAAGAGCGCCAAATAGCGGTGATTTGGGGAATTCAGAAAGAACCAATGCTG 824
QY 221 ProAspPheSerLysAlaIlePheAlaLeuThrProGlyAspTyrThrLysThrProVal 240
DB 825 CCGGATTTTNCCTAAAGCCGCTTCCGCTTAATCTCTGCGGGAATTACATTAAGCCCTGTT 884
QY 241 LysThrGluPheGlyTyrHisIleIleTyrLeuIleSerIysAspSerProValThrTyr 260
DB 885 AAAACAGAGTTGGTTATCATATTTATCTATTTGATTTCTAAAGATAGCCCTGTAACCTTAT 944
QY 261 ThrTyrGluGlnAlaIleProThrIleIleGlyMetLeuGlnGluIlePheGlnGlu 280
DB 945 ACTTATGAGCAAGCTTAACCTACATTAAGGGGATGTTACAGAAAGCTTTCCAGAGA 1004
QY 281 ArgMetAsnGlnArgIleGluGluLeuArgIleHisAlaIleValIleAsnIys 299
DB 1005 CGCATGAATCAACGATTAAGAGATTAAAGAAAGACGCTAAATGTTTCAACAG 1061
RESULT 8
ABX66063 standard; DNA; 844 BP.
ID ABX66063
XX ABX66063;
AC ABX66063;
XX 07-MAY-2003 (first entry)
DT Helicobacter pylori selected interacting domain (SID) DNA #662.
DE Helicobacter pylori selected interacting domain; SID;
KW gene; de.

```

XX OS Helicobacter pylori.
XX PN MO20026501-A2.
XX PD 29-AUG-2002.
XX PF 28-DEC-2001; 2001WO-BP015428.
XX PR 02-JAN-2001; 2001US-0259302P.
XX PA (HYBR-) HYBRIGENICS.
XX PA (INSP) INST PASTEUR.
XX PI Legrain P, Rain J, Collard F, De Reuse H, Labigne A;
XX DR WPI: 2002-674910/72.
XX DR P-PSDB; ABU51319.
XX PT New complexes of protein-protein interactions in Helicobacter pylori,
XX PT useful for identifying modulating compounds for treating or preventing
XX PT ulcers in mammals.
XX PS Claim 7, Page 245; 642pp; English.
XX CC The invention describes a complex of protein-protein interactions in
XX CC Helicobacter pylori selected from 421 complexes given in the
XX CC specification. The complex of protein-protein interactions are useful for
XX CC screening for agents which modulate the interaction of proteins.
XX CC Modulating compounds which binds to a targeted bacterial protein may be
XX CC used for treating or preventing ulcers in a human or animal. This
XX CC sequence encodes a selected interacting domain (SID), identified via
XX CC protein-protein interactions
XX SQ Sequence 844 BP; 315 A; 143 C; 175 G; 211 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9, 23e-120 Length: 844
Score: 1311.00 Matches: 257
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatch: 0
Query Match: 86.25% Indels: 0
DB: Gaps: 0

US-10-039-183a-2 (1-239) x ABX66063 (1-844)
QY 43 GYVALLLEuAlaThValAspGlyArgProIleThrIysSerAspPheAspMetIleIys 62
Db 3 GCGGTGTTAGCGACAGTGGATGCGACCTATCACTAAAGCGATTGACATGATTAAG 62
QY 63 GlnATGAenProAsnPheAspPheAspIleuIysGluIysGluIleuIle 82
Db 63 CAACGAATTCCTAATTTTTCATTTTGCACAGCTTAAAGAGAAAGAAAGAACCTTTGATT 122
QY 83 AspGlnAlaIleArgThrAlaLeuValGluAengIuAlaIysThrGluIysLeuAspSer 102
Db 123 GATCAAGCATATTGCGACCGCCTTGTAGAAATGAAAGCAAGCAAGCAAGC 182
QY 103 ThrProGluPheIleAlaMetMetGluAlaValIysGluAlaLeuValGluIleuThr 122
Db 183 ACTCCAGAAATTTTAAAGCGATGATGAGAGCGGTTAAAAACAGGCTTTAGTGAATTTTGG 242
QY 123 AlaIysIysGluAlaGluIuValIysIysValGluIleProGluIysGluMetGluAsp 142
Db 243 GCTAAATAAAACAGCGCTGAAGAACTGAAAAAGTCCAAATCCGAAAAAGAAATGCAAGAT 302
QY 143 PheIysAsnAlaAsnIysAspGluIleuPheValIysGluIleAlaIleArgHisIle 162
Db 303 TTTTCAACGCTAAACAAAGATCAGCTTTTGTCAAGCAAGAGCCCATGCTAGCATATT 362
QY 163 LeuValIysThrGluAspGluAlaIysArgIleIleSerGluIleAspIysGluProIys 182
Db 363 TTATGAAAAACGAAAGATGAGGCTAAACGATTAATTTTCGAGATTGACAAACAGCCAAAG 422

```

```

QY 183 AlaIysIysGluAlaIysPheIleGluIleuAlaAsnArgAspThrIleAspProAsnSer 202
Db 423 GCTAAATAAAAGAAAGTCAATTCATTAAGTACCGCAATCGGATACATGATGATCTAAACGC 482
QY 203 LysAsnAlaGlnAsnArgIleValAspIleuGluIysPheGluIysAsnGluMetAlaProAsp 222
Db 483 AAGAACGGGCACAAAGAGGGGATTTGGGAAATTTCCAAAGAACCAATGAGCTCCGAT 542
QY 223 PheSerIysAlaAlaPheAlaLeuThrProGluIysAspIysThrIleThrIysThr 242
Db 543 TTTTCTAAAGCCGCTTTGCTTTAACTCTCGGAGATTACACTAAACCCCTGTTAAACA 602
QY 243 GluPheGluIysIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 262
Db 603 GAGTTGGTTATCATATTATCATATTGATTTCTTAAAGATAGCCCTGTAACCTAATCTAT 662
QY 263 GluGlnAlaIysProThrIleIysGluIleuIleuIleuIleuIleuIleuIleuIleuIleu 282
Db 663 GAACAGGCTTAACCTACCATTAAGGGGATGTTTACAAAGAAAGCTTTTCCAGAACGCAAG 722
QY 283 AsnGlnArgIleGluIleuArgIysIleAlaIysIleValIleAsnIys 299
Db 723 AATCAACCATTTGAGCAACTAAGAAAGCAAGCTTAATTTGTTTCAACAG 773

RESULT 9
AAV90546
ID AAV90546 standard; DNA; 616 BP.
XX AC AAV90546;
XX DT 20-MAR-2003 (revised)
XX DT 18-FEB-1999 (first entry)
XX DE Nucleotide sequence from clone D7.
XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX OS Helicobacter pylori.
XX PN MO9849314-A2.
XX PD 05-NOV-1998.
XX PF 25-APR-1998; 98MO-US008487.
XX PR 25-APR-1997; 97US-0045107P.
XX PR 14-OCT-1997; 97US-0061958P.
XX PA (GENE-) GENELABS TECHNOLOGIES INC.
XX PI Chow TP, Fly KE, Lim MY, Mcatee CP;
XX DR WPI: 1999-009433/01.
XX DR P-PSDB; AAM89814.
XX PT New Helicobacter pylori antigens and related nucleic acid sequences -
XX PT useful in serological diagnosis and protective vaccines, providing long-
XX PT lasting immune response.
XX PS Claim 27; Page 97; 402pp; English.
XX CC The present sequence encodes a Helicobacter pylori antigenic protein that
XX CC is characterised by immunoreactivity with H. pylori-positive antisera.
XX CC The proteins are highly immunogenic and induce a long-lasting immune
XX CC response that persists even after antimicrobial treatment. In antibody-
XX CC detection assays, on sera, plasma, urine, saliva etc., they are highly
XX CC sensitive and specific. The specification also describes 69 previously
XX CC unrecognised immunogenic cluster families. H. pylori antigens are used to
XX CC detect H. pylori-specific antibodies, for diagnosing infection or to
XX CC confirm eradication of infection, and in vaccines to protect against H.
XX CC pylori infection and related diseases (gastritis, peptic ulcer, gastric

```



```

QY 41 SerAlaGlyValIleuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMet 60
Db 136 TCAGCAGGCGGTGTTGGCAGCAGTGGATGGCAGCCCATCCAAAAGCATTGATATG 195
QY 61 ILeYSGINARAsnProAsnPheAspPheAspLysLeuLysGluLysGluVal 80
Db 196 ATTAAGCAGCAAGAAATCTCAATTGTTGTTGACAGCTTAAAGAGAAAAGAAAGAGCC 255
QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGlnAlaLysThrGluLysLeu 100
Db 256 TTGATTTGACAGCAAGCTATCCGCAACCGCTTGTGTAATAAGAGCTAAGCAGAAAGCTT 315
QY 101 AspSerThrProGluPheLysAlaMetMetGluAlaValLysGlnAlaLeuValGlu 120
Db 316 AATCAGACTCCAGAAATTTAAAGCATGATGAAAGCGCTTAAAAACAGGCTTTAGTGGAA 375
QY 121 PheTTPAlaLysLysGlnAlaGluGluVal-LysLysValGlnIleProGluLysGlu 140
Db 376 TTTTGGGCTTAAAAACAGGCTGAGAAAGTGAAGAAAGATCCCAATCCAGAAAAGAAAT 435
QY 140 LGINAspPheTYrAsnAla 146
Db 436 GCAGGATTTTTCACACGCC 454

RESULT 11
ABX66722
ID ABX66722 standard; DNA; 595 BP.
AC ABX66722;
XX
XX 07-MAY-2003 (first entry)
DT
DE Helicobacter pylori selected interacting domain (SID) DNA #1321.
XX
XX Protein-protein interaction; ulcer; selected interacting domain; SID;
KM gene; ds.
XX
XX Helicobacter pylori.
OS
XX
XX WO200266501-A2.
PN
PD 29-AUG-2002.
XX
XX 28-DEC-2001; 2001WO-EP015428.
PF
XX 02-JAN-2001; 2001US-0259302P.
PR
XX (HYBR-) HYBRIGENICS.
PA (INSP) INST PASTEUR.
XX
XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
PI
XX
XX WPI; 2002-674910/72.
DR P-PSDB; ABUS1978.
XX
XX New complexes of protein-protein interactions in Helicobacter pylori,
PT useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals.
XX
XX Claim 7, Page 405; 642p; English.
PS
XX
XX The invention describes a complex of protein-protein interactions in
CC Helicobacter pylori selected from 421 complexes given in the
CC specification. The complex of protein-protein interactions are useful for
CC screening for agents which modulate the interaction of proteins.
CC Modulating compounds which binds to a targeted bacterial protein may be
CC used for treating or preventing ulcers in a human or animal. This
CC sequence encodes a selected interacting domain (SID), identified via
CC protein-protein interactions
XX
XX Sequence 595 BP; 205 A; 97 C; 130 G; 163 T; 0 U; 0 Other;
SQ

```

Alignment Scores:

```

Pred. No.: 3.26e-56 Length: 595
Score: 664.00 Matches: 129
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.68% Indels: 0
DB: 6 Gaps: 0
US-10-039-183a-2 (1-299) x ABX66722 (1-595)

QY 121 LysArgIleIleSerGluIleAspLysGlnProLysAlaLysLysGluAlaLysPheIle 190
Db 3 AAACGATTAATTTCTGAGATTGACAAACAGCCAAAGGCTTAAAAAGAAAGCTAAATTCATT 62
QY 191 GluLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyGlyAsp 210
Db 63 GAGTTAGCAATCGGGATACGATTGATCTTAACAGCAAGAAACGGCAAAATGGCGGTGAT 122
QY 211 LeuGlyLysPheGlnLysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeu 230
Db 123 TTGGGGAATTCCAAAAGAACCAATGGCTCCGATTTTCTAAAGCCGCTTTCCTTTAA 182
QY 231 ThrProGlyAspTYrThrLysThrProValLysThrGluPheGlyTYrHisIleTYr 250
Db 183 ACTCCTGGGAGATTACACTPAAACCCCTGTAAACAGAGTTGGTTATCATATTATCTAT 242
QY 251 LeuIleSerLysAspSerProValThrTYrThrTYrGluGlnAlaLysProThrIleLys 270
Db 243 TTGATTTCTAAAGATAGCCCTGTAACTTAATCTTAATGAACAGGCTAAACCTAACCTTAAG 302
QY 271 GlyMetLeuGlnGluLysLeuPheGlnGluArgMetLeuAsnGlnArgIleGluGluLeuArg 290
Db 303 GGGATGTTTACAAAGAAAGCTTTTCCAAAGACGATGATCAACGATTAAGGAACTAAGA 362
QY 291 LysHisAlaLysIleValIleAsnLys 299
Db 363 AAGCAGCTTAAATTTTATCAACAG 389

RESULT 12
AAQ22604
ID AAQ22604 standard; DNA; 960 BP.
XX
XX AAQ22604;
AC
XX
XX 25-MAR-2003 (revised)
DT
DT 22-JUL-1992 (first entry)
XX
XX Gene encoding OMP of Bordetella pertussis.
DE
XX
XX Outer membrane protein; vaccine; whooping cough; haemophilus; influenza;
KM hepatitis; herpes; ss.
XX
XX Bordetella pertussis.
OS
XX
XX Key Location/Qualifiers
FH 101..874
FT CDS /*tag= a
XX
XX EP476287-A.
XX
XX 25-MAR-1992.
PD
XX
XX 03-AUG-1991; 91EP-00113094.
PF
XX 21-SEP-1990; 90US-00586347.
PR
XX (AMCY) AMERICAN CYANAMID CO.
XX
XX ReillyPA, Reilly PA;
PI
XX
XX WPI; 1992-098085/13.
DR P-PSDB; AAR21691.
XX
XX Nucleotide andaminoacid sequence of protein-B-pertussis - useful as
XX

```



```

QY 5 ILeuAAsnLeuAlaLeuValGlyAlaLeuSerThrPheLeuMetAlaLysProAla 24
   |||||
DB 22 TTAAGTAATAAATATATCTTACCTGTAACAGCAAGTGCCTTTGTAGTCCCTGTGT 81
   |||||
QY 25 HisAAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrThrAspSerSerAlaGlyVal 44
   |||||
DB 82 TCAATATGCAACTGATTCGAAAGATTAACA-----CTTATTTCTTCTTAAGCTGTGAT 135
   |||||
QY 45 LeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMetIleLysGlnArg 64
   |||||
DB 136 GTTAAAGTTGCTGAC-----GTAATGAAAAA-----ATGGCAAGAAAGCA 177
   |||||
QY 65 AsnProAsnPheAspPhe-----AspLysLeuLysGlu 75
   |||||
DB 178 ATTGCTAACACACTCTTTTACGATGTGATTAACAAAGTTTACGAGTAAATTAAGAT 237
   |||||
QY 76 LysGluLysGluAlaLeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGluAla 95
   |||||
DB 238 AAAAGTAGATACAAAGATATGATTAAGATTTAAAAAGAGAAACAATATACGCGGT 297
   |||||
QY 96 LysThrGlu-----LysLeuAspSerThrProGlu 105
   |||||
DB 298 AAAAGTCAATTTGAAAGCATGTAAAAACAAGGTATGTCATTTGATTAATAAGAA 357
   |||||
QY 106 PheLysAlaMetMetGluAlaValLysLysGlnAlaLeuValGluPheThrAlaLysLys 125
   |||||
DB 358 CAGAAG---AAATTATCAGCTTATCAGAAACAATTTCTTCTGAT----- 399
   |||||
QY 126 GlnAlaGluGluValLysLysValGlnIleProGluLysGluMetGlnAspPheThrAsn 145
   |||||
DB 400 -----AAAGTTAATGATCATCAATAAAGAAATCAAAAGAA-----AAC 435
   |||||
QY 146 AlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHisIleLeuVal--- 164
   |||||
DB 436 TCTAAAAA-----ACATCACACATTTTAATTA 465
   |||||
QY 165 -----LysThrGluAspGluAla 170
   |||||
DB 466 GTTAATATCTAATCTAGCGATTAAGAGTTTATCATGATTAATAAAGAAAGAAAGT 525
   |||||
QY 171 LysArgIleIleSerGluIleAspLysGlnProLysAlaLysLysGluAlaLysPheIle 190
   |||||
DB 526 GAAAGAAATTCAAAAAGATGAAAAAGATCCAAAT-----AAGTTTGA 570
   |||||
QY 191 GlnLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyGlyAsp 210
   |||||
DB 571 GAAATATGCAAAAAAAGAAATCAATGAC-----AGTTCTTCAGCTAAGAAAGATGCTCA 624
   |||||
QY 211 LeuGlyLysPheGlnLysAsnGlnMetAlaProAspPheSerLysAlaIleAspAlaLeu 230
   |||||
DB 625 TTAGTTATGATTAATTAAGCTCAATGATGATAGCTTGAAGAAAGCACTATTAAATTA 684
   |||||
QY 231 ThrProGlyAspThrThrLysThrProValLysThrGluPheGlyThrHisIleIleThr 250
   |||||
DB 685 AAAAGAGGAGAGATTCTTAAGTC---GTAAGAAACAGACTATGCGTATCACTT----- 735
   |||||
QY 251 LeuIleSerLysAspSerProValThrThrLysThrGluAlaLysProThrIleLys 270
   |||||
DB 736 ---ATTAAGACAGATTAAGAAACGATTTTAATAGTAA-----AAA 774
   |||||
QY 271 GlyMetLeuGlnGluLysLeuPheGlnGluArgMetAsnGlnArg----- 285
   |||||
DB 775 TCAATATATCAAAACAAATTAATTCAGAGAAAGATGACAGAAAGAAACCTTAATTA 834
   |||||
QY 286 -----IleGluGluLeuArgLysHisAlaLysIle 295
   |||||
DB 835 GATGCATACAAAGAAATTTAATAAGGATTAATAAGTA 870
   |||||
RESULT 14
ADSO2896
ID ADSO2896 standard; DNA; 993 BP.
XX

```

```

AC ADSO2896;
XX
XX 04-NOV-2004 (first entry)
DT
XX
XX Staphylococcus epidermidis polynucleotide seqid 2191.
DE
XX
XX Staphylococcus epidermidis vaccine; antisense therapy; Staphylococcus epidermidis;
XX antibacterial; recombinant expression vector; infection; computer readable medium;
XX recombinant expression vector; infection; computer readable medium;
XX computer based system; gene; ds.
XX
XX Staphylococcus epidermidis.
XX
XX US200414734-A1.
XX
XX 29-JUL-2004.
XX
XX 01-DEC-2003; 2003US-00724972.
XX
XX 08-NOV-1997; 97US-0064964P.
XX 13-AUG-1998; 98US-00134001.
XX 29-NOV-1999; 99US-00450969.
XX
XX (DOUC/) DOUCETTE-STAMM L.
XX (BUSH/) BUSH D.
XX
XX Doucette-Stamm L, Bush D;
XX
XX MPI: 2004-580138/56.
XX P-PSDB: ADS06668.
XX
XX New isolated polypeptide and encoding nucleic acid derived from
XX Staphylococcus epidermidis, useful for diagnosing, preventing and/or
XX treating an S. epidermidis bacterial infection.
XX
XX Claim 5; SEQ ID NO 2191, 741pp: English.
XX
XX The invention describes an isolated nucleic acid comprising a nucleotide
XX sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
XX 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
XX of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
XX given in the specification. Also described are: a recombinant expression
XX vector; a cell comprising a recombinant expression vector of (1);
XX CC producing an S. epidermidis polypeptide; an isolated nucleic acid
XX comprising a nucleotide sequence of at least 8 nucleotides in length; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection, comprising a nucleic acid cited above and a carrier; treating
XX a subject for S. epidermidis infection; a recombinant or substantially
XX pure preparation of an S. epidermidis polypeptide or its fragment; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection; detecting the presence of a Staphylococcus nucleic acid in a
XX sample; a computer readable medium having recorded in it the nucleotide
XX sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
XX system for identifying fragments of the Staphylococcus genome of
XX commercial importance; a computer based system for identifying fragments
XX of the Staphylococcus plasmids of commercial importance; identifying
XX commercially important nucleic acid fragments of the Staphylococcus
XX genome and/or plasmids; and identifying an expression modulating fragment
XX of the Staphylococcus genome and/or plasmid. The methods and
XX compositions of the present invention are useful for the diagnosis,
XX prevention and/or treatment of an Staphylococcus epidermidis bacterial
XX infection. This sequence encodes a S. epidermidis protein of the invention.
XX
XX Sequence 993 BP; 458 A; 117 C; 166 G; 252 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1.31e-12 Length: 993
XX Score: 224.00 Matches: 85
XX Percent Similarity: 42.77% Conserv: 57
XX Best Local Similarity: 25.60% Mismatches: 100
XX Query Match: 14.74% Indels: 90
XX DB: 13 Gaps: 16
XX
XX US-10-039-183A-2 (1-239) x ADSO2896 (1-993)

```

```

QY 5 ILeuAenLeuAlaLeuValGlyAlaLeuSerThSerPheLeuMetAlaLysProAla 24
DB 22 TTAATGAAATTAATTAATCTTACCTGTAACAGCAAGCACTTTGTAGTGCCTGTGT 81
QY 25 HLeuAenAlaLeuAenAlaThrhIsaenThrLysLysThrThraspSerSerAlaGlyVal 44
DB 82 TCAAAATGCAATGCAATGCAAGGATACACA-----CTTATTCTTCTTAAGCTGTGTAT 135
QY 45 LeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMetIleLysGlnArg 64
DB 136 GTTAAAGTTCCTGAC-----GTAATGAAAAA-----ATGGCAAAAGAACAA 177
QY 65 AsnProAsnPheAspPhe-----AspLysLeuLysGlu 75
DB 178 ATTGCTACACACTCTTTAGCATGTATTAACAAAGTTTACAGATTAATTAATAGAT 237
QY 76 LysGluLysGluAlaLeuIleAspGlnAlaIleArgThraLeuValGluAsnGluAla 95
DB 238 AAAAGTAGATACAAAGATATGTATTAAGATATTAAGAAAGAAAGAAACATACGCGGT 297
QY 96 LysThrGlu-----LysLeuAspSerThrProGlu 105
DB 298 AAAAGATCAATTTGAAAGCATGTAAACAAACAGATATGCTCATGATGATTATAAGAA 357
QY 106 PheLysAlaMetMetGluAlaValLysLysGlnAlaLeuValGluPheThrAlaLysLys 125
DB 358 CAGAAAG---AAATATACACTTATGCAAAACATTTGTTACTGAT----- 399
QY 126 GlnAlaGluGluValLysLysValGlnIleProGluLysGluMetGlnAspPheTyraen 145
DB 400 -----AAAGTTAATGTATCAGATTAAGAAATCAAAAGAA-----AAC 435
QY 146 AlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHisIleLeuVal--- 164
DB 436 TCTAAAGAA-----ACATCAACACTTTTAATTAATA 465
QY 165 -----LysThrGluAspGluAla 170
DB 466 GTTAAATCTAAATCTTACGATTAAGAAAGTTTATCATGATTAAGAAAGAAAGAAAGCT 525
QY 171 LysArgIleIleSerGluIleAspLysGlnProLysAlaLysGlnAlaLysPheIle 190
DB 526 GAAAAAATTCAAAAAGAAAGTGGAAAAAGATCCAAAT-----AAGTTTGA 570
QY 191 GlnLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyLysAsp 210
DB 571 GAAATAGCAAAAAAAGAAATCAATGAGC-----AGTTCTTCACTTAAGAAAGATGCTTCA 624
QY 211 LeuGlyLysPheGlnLysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeu 230
DB 625 TTAGGTTATGTATTAAGCTCAAAATGTAGATAGCTTGAAGAAAGCACTTATTAATTA 684
QY 231 ThrProGlyAspTyThrLysThrProValLysThrGluPheGlyTyThrHisIleIleTy 250
DB 685 AAAAGAGGAGAGCTTCTAAAGC---GTAABAAACAGACTATAGCGCATACATT----- 735
QY 251 LeuIleSerLysAspSerProValThrTyThrTyThrGluGlnAlaLysProThrIleLys 270
DB 736 ---ATTAAGACAGATTAAGAAAGCTGATTTTAATAGGAA-----AAA 774
QY 271 GlyMetLeuGlnLysLeuPheGlnGlnArgMetAsnGlnArg----- 285
DB 775 TCAAAATATCAAAACAAAATTAATCGAAGAAAGGTACAGAGAAACCTTAATTAATTA 834
QY 286 -----IleGluGluLeuArgLysHisAlaLysIle 295
DB 835 GATGCATACAAAGAAATTAATTAAGAGCTTAAGAAAGTA 870

```

```

AC ABO70541;
XX
DT 29-AUG-2003 (revised)
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #483.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes ATCC 19115.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR003061.
XX
PR 04-OCT-2000; 2000FR-00012697.
XX
PA (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaeser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
XX
PS Claim 14; SEQ ID NO 3354; 180bp; French.
XX
CC The present invention relates to nucleic acid sequences (AB067188-
CC AB01212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from Wipo at ftp.wipo.int/pub/published_pct_sequences. (updated
CC on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 1720 BP; 500 A; 299 C; 320 G; 601 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 4,466-12 Length: 1720
Score: 221.50 Matches: 83
Percent Similarity: 43.31% Conservative: 53
Best Local Similarity: 26.43% Mismatches: 113
Query Match: 14.57% Indels: 65
DB: 6 Gaps: 15
XX
US-10-039-183a-2 (1-299) x ABO70541 (1-1720)
QY 2 LysLysAsnIleLeuAsnLeuValGlyAlaLeuSerThSerPheLeuMetAla 21
DB 1354 AAGAAATTAATCTTGAAGCTTGTCAATGATGAGGCAATG-----TTCAAGTCTAGCA 1304
QY 22 LysProAlaHisAsnAlaAsnAlaThrhIsaenThrLysLysThrThraspSerSer 41
DB 1303 -----GCCGCGCGTGC 1292
QY 42 AlaGlyValLeuAlaThrValAspGlyArgProIleThrLysSerAsp---PheAspMet 60
DB 1291 GGTGAGATGTCTTAAGACAGACTCTGCGCATGTACAGCAAGATGAGCTTTACAGCGCA 1232
QY 61 IleLysGlnArgAsn-----ProAsnPheAspPheAspLysLeu----- 73
DB 1231 ATGAAAGATTAATATGTTCTGATGTTGTGACAAACATTAATCTTGAAGAAATTCCTGGA 1172

```

RESULT 15
 ABO70541/c
 ID ABO70541 standard; DNA; 1720 BP.
 XX

```
QY 74 ---LysGluLysGluLysGluAlaLeuIleAspGlnAlaIleArgThrAlaLeuValGlu 92
DB 1171 GATTAATACAAAGTAAGCAGACGAGTGTGATAGAAATTC----- 1130
QY 93 AengGluAlaLysThrGluLysLeuAspSerThrProGluPheLysAlaMetMet----- 110
DB 1129 AACGAGTAAATCAACAATACGAGAT-----CAATCTCTGCACTTTAGCTCAA 1079
QY 111 -----GluAlaValLysLysGlnAlaLeuValGluPheTrpAlaLysLys 125
DB 1078 AGTGCTTAACGAGAAATCAATCAAAAGCCAACTTAAGTACAACTTTAGTCAAAA 1019
QY 126 GlnAlaGluLysValLysValGlnIleProGluLysGluMetGlnAspPheTyrAsn 145
DB 1018 GGTACTGAA-----GCCAATACTGATACTAGCAGACAAACATTAATAAATACTATGAA 965
QY 146 AlaAsnLysAspGlnLeuPheValLysGlnAlaIleAlaArgHisIleLeuValLys 165
DB 964 ACATGGCAACCAAGATATTACTGTAAGC-----CATATTCCTTGAGCT 923
QY 166 ThrGluAspGluAlaLysArgIleIleSerGluIleAspLysGlnProLysAlaLysLys 185
DB 922 GATGAAACAAAGCCAAAGAAAGT-----GAACAAAAACTGAAAGAC 881
QY 186 GlnAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAla 205
DB 880 GCGCGAAATTTGCTGATTATAGCAAAAGAAATACTACAGACACTGCTACTAAAGAC--- 824
QY 206 GlnAsnGlyGlyAspLeuGlyLysPheGlnLysAsnGlnMetAlaProAspPheSerLys 225
DB 823 ---AATGTGGACAATTAAGTCCATTCGCTCGTGAAGATGATCTCATTTTGAAAAA 767
QY 226 AlaAlaPheAlaLeuThrProGlyAspTyrThrLysThrProValLysThrGluPheGly 245
DB 766 GCAGCTTATGCTCTTAATAAAACAAAGCGACATCAGCGCTCCAGTAAAAACACATACGGA 707
QY 246 TyrHisIleIleTyrIleuIleSerLysAspSerProValThr---TyrThrTyrGluGln 264
DB 706 TACCAATCATCAATG-----GACAAACCTGCACAAACAAAAACATTCGAAAAA 656
QY 265 AlaLysProThrIleLys---GlyMetLeuGlnGluLysLeuPheGlnGluArgMetAsn 283
DB 655 GATAAAAAAGCTGTAAGCTTCTTAAGTCTCACTAACTGAGAACATGCAA 596
QY 284 GlnArgIleGluGluLeuArgLysHisAlaLysIleValIle 297
DB 595 AAAACGCTTAATAAAAGAAATACAAAGATGCTAACGTAAAGTG 554
```

Search completed: April 24, 2005, 06:13:28
Job time : 550.741 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: Apr11 24, 2005, 05:33:37 ; Search time 171.775 Seconds
(without alignment)
2848.182 Million cell updates/sec

Title: US-10-039-183A-2
Perfect score: 1520
Sequence: 1 MKKQILNALVGLSTSLFLM.....ERMNQRIEELRKAKIVINK 299

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+g2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10039183/runat_22042005_122426_20141/app_query.fasta_1.1038
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=ext1 -MINMATCH=0.1 -LOOFC=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=humand40.cdi
-LIST=445 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10039183@cgn2_1_219@runat_22042005_122426_20141 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLIT=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCUTS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1520	100.0	1149	4	US-09-336-115C-23
2	253.5	16.7	2032	4	US-09-902-540-8077
3	253.5	16.7	7811	4	US-09-902-540-824
4	224	14.7	993	3	US-09-134-001C-974
5	211.5	13.9	2097	4	US-09-252-991A-13635
6	211.5	13.9	2331	4	US-09-252-991A-13809
7	211.5	13.9	63563	4	US-09-596-002-33
8	206.5	13.6	7963	4	US-08-956-171E-168
9	206.5	13.6	7963	4	US-08-781-986A-168
10	205.5	13.5	1110	4	US-09-540-236-1818
11	199.5	13.1	1371	4	US-09-902-540-3927
12	199.5	13.1	21375	4	US-09-902-540-1193

13	180.5	11.9	1830121	4	US-09-557-884-1	Sequence 1, Appl
14	180.5	11.9	1830121	4	US-09-643-990A-1	Sequence 1, Appl
15	174	11.4	825	4	US-09-252-991A-1818	Sequence 3818, Ap
16	174	11.4	855	4	US-09-252-991A-3642	Sequence 3642, Ap
17	174	11.4	1785	4	US-09-252-991A-3774	Sequence 3774, Ap
18	174	11.4	1869	4	US-09-252-991A-3694	Sequence 3694, Ap
19	172	11.3	1332	4	US-09-328-352-3599	Sequence 3599, Ap
20	171	11.2	1224	4	US-09-252-991A-1865	Sequence 3865, Ap
21	156.5	10.3	1005	3	US-09-277-565-14	Sequence 14, Appl
22	156.5	10.3	1830121	4	US-09-557-884-1	Sequence 1, Appl
23	156.5	10.3	1830121	4	US-09-643-990A-1	Sequence 1, Appl
24	152.5	10.0	1296	4	US-09-252-991A-818	Sequence 818, Ap
25	152.5	10.0	1344	4	US-09-252-991A-851	Sequence 851, Ap
26	150.5	9.9	1113	4	US-09-252-991A-12902	Sequence 12902, A
27	150.5	9.9	1374	4	US-09-252-991A-13227	Sequence 13227, A
28	150	9.9	333	4	US-09-489-038A-859	Sequence 859, Ap
29	146.5	9.6	1416	4	US-09-252-991A-833	Sequence 833, Ap
30	144	9.5	318	4	US-09-489-038A-844	Sequence 844, Ap
31	144	9.5	336	4	US-09-252-991A-3451	Sequence 3451, Ap
32	144	9.5	1218	4	US-09-252-991A-3440	Sequence 3440, Ap
33	142.5	9.4	279	4	US-09-408-020-67	Sequence 67, Appl
34	142.5	9.4	1341	4	US-09-489-038A-5112	Sequence 511, Ap
35	142.5	9.4	3298	4	US-09-408-020-1	Sequence 1, Appl
36	139	9.1	294	4	US-09-252-991A-3921	Sequence 3921, Ap
37	134.5	8.8	42432	4	US-09-408-020-2	Sequence 2, Appl
38	131	8.6	1881	4	US-09-328-352-3710	Sequence 3710, Ap
39	129	8.5	321	4	US-09-328-352-1085	Sequence 1085, Ap
40	128.5	8.5	105733	4	US-09-949-016-11080	Sequence 13080, A
41	128	8.4	1764	4	US-09-489-038A-6011	Sequence 6011, Ap
42	127	8.4	62909	4	US-09-596-002-32	Sequence 32, Appl
43	126.5	8.3	497	4	US-09-621-976-3178	Sequence 3178, Ap
44	126.5	8.3	605	3	US-09-208-804-2	Sequence 2, Appl
45	126.5	8.3	605	3	US-08-801-743-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-336-115C-23
Sequence 23, Application US/09336115C
Patent No. 6576244
GENERAL INFORMATION:
APPLICANT: Wetzlin, Richard A.
APPLICANT: Guy, Bruno
TITLE OF INVENTION: LT and CT in Parenteral Immunization
TITLE OF INVENTION: Methods Against Helicobacter Infection
FILE REFERENCE: 06132/055002
CURRENT APPLICATION NUMBER: US/09/336,115C
CURRENT FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: US 09/100,258
PRIOR FILING DATE: 1998-06-19
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 1149
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (106)...(1002)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: (106)...(1166)
US-09-336-115C-23

Alignment Scores:

Pred. No.: 3,836-166
Score: 1520.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4
Length: 1149
Matches: 299
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-039-183a-2 (1-299) x US-09-336-115C-23 (1-1149)

QY 1 MetLysLysAsn11LeuAsn1LeuVal1G1Val1LeuSerThrsPheMet 20
106 ATGAAAAAATATCTTAATTTAGGTGAGTGGGCGCTGAGCGCGTGTGATG 165

QY 21 AlaLysProAlaHisAsn11AsnAsn1aThrHisAsnThrsLysThrsPheSer 40
166 GCTAAGCCGGCTCATTAAGCAATAATACGCTACGATTAACAGAAAAAAGCATGTTCT 225

QY 41 Ser1aG1Val1Leu1aThrVal1aSpG1YArPro11eThrsLysSerAspPheMet 60
226 TCAGCAGCGGTGTGCGACAGTGCAGACCTATCATCTAAAGCATTTTGCATG 285

QY 61 IleLysG1aThrAsnProAsn1PheAsn1PheAsn1PheLysG1uLysG1uVal1a 80
286 ATTAAAGCAAGAAATCTTAATTTGATTTTGAACAGCTTAAGAGAAAGAAAGAGCC 345

QY 81 Leu1LeaSPG1a11eArGThrs1aLeuVal1G1uAsnG1u1aLysThrsG1uLysLeu 100
346 TTGATTTGATCAAGCTATTGCAACCGCCCTGTAGAAAATGAAGCTAAACCGAAATTG 405

QY 101 AspSerThrsProGluPheLys1aMetMetG1u1aVal1LysG1u1aLeuVal1G1u 120
406 GACACACACTCCAGAAATTTAAAGCATGATGAAAGCGCTTAAAGAAACAGCGCTTAAAGGAA 465

QY 121 PheTP1aLysLysG1u1aG1uG1uVal1LysLysVal1G1u1eProG1uLysG1uMet 140
466 TTTTGGGCTTAAGAAACAGGCTGAGAGATGAAAGTCAAAATCCCAAGAAAGAAATG 525

QY 141 GlnAspPheThrsAsn1aAsnLysAspG1uLysPheVal1LysG1u1aHis1aArg 160
526 CAAGATTTTTCACAGCTAAACAAAGATCAGCTTTTGTCAACAGCAAGAACCCATGTAG 585

QY 161 His1eLeuVal1LysThrsG1uAspG1u1aLysArg1e1eSerG1u1eAspLysG1u 180
586 CATATTTTAAGTAAACCAAGATGAGCTAAACGATTTATTTGAGATTGACAAACAG 645

QY 181 ProLys1aLysLysG1u1aLysPhe1eG1uLeu1aAsnArgAspThrs1eAspPro 200
646 CCAAGGCTTAAGAAAGAACTTAATTCATTGAGTTAGCCAAATCGGATTCGATTCCT 705

QY 201 AsnSerLysAsn1aG1uAsnG1u1aLysPheLysG1uLysPheG1uLysAsnG1uMet1a 220
706 AACACAGAAACGCGCAAAATGCGGTGATTTGGGGAATTCACAAAGAACCAATGCT 765

QY 221 ProAspPheSerLys1a1aPheAlaLeuThrsProG1YAspThrsThrsThrsProVal 240
766 CCGGATTTTCTTAAGCCGCTTTCCTTAACTCCCTGGGATTTACCTAAACCCCTGTT 825

QY 241 LysThrsGluPheG1YThrs1e1eThrsLys1eSerLysAspSerProValThrsThrs 260
826 AAAACAGATTTGTTATATATATCTATTTGATTTCTTAAGATAGCCCTGTAACTTAT 885

QY 261 ThrYrG1uG1u1aLysProThrs1eLysG1uMetLeuG1uLysLeuPheG1u1a 280
886 ACTTATGAACAGGCTAAACCTACCTAATTAAGGGAATTTACAGAAAGCTTTTCCAGAA 945

QY 281 ArgMetAsnG1aArg1eG1uG1uLeuArg1eSH1aLys1eVal1LeuAsnLys 299
946 CGCATGAATCAACGATTAAGAGAACTTAAGAAAGCAAGCTTAAATTTTATCAACAAAG 1002

Db

RESULT 2
US-09-902-540-8077
; Sequence 8077, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B

US-10-039-183a-2 (1-299) x US-09-902-540-8077 (1-2032)

QY 22 LysProAlaHisAsn1aAsnAsn1aThrHisAsnThrsLysThrsPheSer- 41
45 GCGCTGCGCGCGGGGTGGCGATCCGCTGCACCAAGCCGATGAGAGAGCGCGGA 104

QY 41 r1aG1YVal1Leu1aThrVal1aSpG1YArPro11eThrsLysSerAspPheAspMet- 60
105 GCGCTGATGTGTGCGCACGCTCAACGCGAGGTCTCAGCGCGGCGGATTCGAACAGGA 164

QY 61 -----1LeysG1aThrAsnProAsn1PheAsn1PheAs 71
165 GCTGTGCGCGAGCTGCGCTGACCGAGCTGCTCCAGCGCAC----- 207

QY 71 p1eLysG1uLysG1u-----LysG1u1aLeu1eAspG1u1a1eArgThrs1a 89
208 ---TTGAGAGAGCTGAGACCTTCAAGCGCGCTCTCGATACGTACATCACCGAT 263

QY 89 AleuValG1uAsnG1u1aLysThrsG1uLysLeuAspSerThrsProGlu----- 105
264 GCTGTGCTGCAGAGAGGCGCGCACAGCAAGTCAACCGTCAACCGGAGAGGTGACCG 323

QY 106 -----PheLys----- 107
324 GGGGCTGCTGCGCGCTGCGGAGACTATCCGCGGCAACTCAACAGTCTGAGCCCA 383

QY 108 -----AlaMetMetG1u1aVal1LysG1u1aLeuVal1GluPheTP1aLys 124
384 GGGGAGCTCTTCATGCGGAGCTGCGCTCGCGGAGAGCAAGCCGCTGACCATGCAAA 443

QY 124 e1eLysG1u1aG1uG1uVal1-----LysLysVal1G1u1eProG1uLysG1uMetG1aAspPh 143
444 GCTGTTCGCCAGCACGCTTATTCGCGGTGCGGTGAGAGAAAGAACTCGGGGCTG 503

QY 143 e1eThrsAsn1aAsnLysAspG1uLysPheVal1LysG1u1aHis1a1aArgHis1eLe 163
504 GTACCGCGCGCAGAGAGACTTCCAGAGCCCGAGAGGTCCAGCGCGCGCAATGCT 563

QY 163 uValLysThrsG1uAspG1u1aLysArg1e1eSerG1u1eAspLysG1uProLys1a 183
564 GGTGAAGGCGCTGCGAGGCGAGCGGCTG-----CAAAACCAAGCT 605

QY 183 AleLysG1u1aLysPhe1eG1uLeu1aAsnArgAspThrs1eAspProAsnSerLys 203
606 CAAGTCCGCAAGAGATTCAGAGACTGCGCGGCGAGTATTCGCT-----AGGCG 656

QY 203 sAsn1aG1uAsnG1Y1aLysPheLysG1uLysPheG1uLysAsnG1uMet1aProAspPh 223
657 GGAAGCGCAAGGTGAGAGGAGCTGGGTCTTCCCGGGGAGATGCGCGCGCTT 716

QY 223 eSerLys1a1aPheAlaLeuThrsProG1YAspThrsThrsThrsProValLysThrsG1 243
717 CGACGAAGTGTATTCAAGCTGGGGGTGGGAGGATTCGAGCGT---GTTCACACGA 773


```

Db      82 TCMAATGCAACTGAAATCGAAGATAACACA-----CTTATTTCTTCTAAGCTGGTGTAT 135
QY      45 LeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMetIleLysGlnArg 64
      136 GTTAAAGTTGCTGAC-----GTAATGAAAAA-----ATGGGCAAAAGAACAA 177
QY      65 AsnProAsnPheAspPhe-----AspLysLeuLysGlu 75
      178 ATTGCTAACACACTCTTTTATGATGTATTAACAAAGTTTATGACGATTAATATAAAGAT 237
Db      76 LysGluLysGluAlaLeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGluAla 95
      238 AAAAGTAGATACAAAAGATATTGATTAAGATATTAAAAAGAAAGAAAGAAATACCGCGCT 297
QY      96 LysThrGlu-----LysLeuAspSerThrProGlu 105
      298 AAAAGATCAATTTGAAAGCATGTTTAAACACAAAGCATATGCTCATGATATTATAAAGAA 357
QY      106 PheLysAlaMetMetGluAlaValLysLysGlnAlaLeuValGluPheThrAlaLysLys 125
      358 CAGAAAG---AAATTATCAGCTTATCAGAAACAATTGTTACTGAT----- 399
QY      126 GlnAlaGluGluValLysLysValGlnIleProGluLysGluMetGlnAspPheTyraen 145
      400 -----AAAGTTATGTATCAGATTAAGAAATCAAAAGAA-----AAC 435
QY      146 AlaAsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHisIleLeuVal--- 164
      436 TCTAAAAA-----ACATCACACATTTTAAATTA 465
QY      165 -----LysThrGluAspGluAla 170
      466 GTTAATATCTAATCTAGCATTAAGAAAGGTTTATCAGATTAATAAAGCGAAAGAAAGCT 525
Db      171 LysAlaGlleIleSerGluIleAspLysGlnProLysAlaLysLysGluAlaLysPheIle 190
      526 GAAAAAATTCAAAAAGAAAGTGGAAAAAGATCCAAAT-----AAGTTTGA 570
QY      191 GluLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyLysAsp 210
      571 GAAATATGCAAAAAAAGAAATCAATGAC-----AGTTCTTCAGCTTAAAGAAAGTGTCA 624
Db      211 LeuGlyLysPheGlnLysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeu 230
      625 TTAGGTTATGTATTAAGCTCAAAATGTAGATAGCTTGAAGAAAGCACTATTAAATTA 684
QY      231 ThrProGlyAspTyThrLysThrProValLysThrGluPheGlyTyThrIleIleTy 250
      685 AAAAGAGGAGAGATTCTTAAAGTC--GTAAAAACAGACTATGGCTATCACT----- 735
Db      251 LeuIleSerLysAspSerProValThrTyThrTyGluGlnAlaLysProThrIleLys 270
      736 ---ATTAAGCAAGATTAAGAAACTGATTTTAATAGTGA-----AAA 774
QY      271 GlyMetLeuGlnGluLysLeuPheGlnGluArgMetAsnGlnArg----- 285
      775 TCAAAATATCAAAACAAATTAATCGAAGAAAGTACAGAGAAACCTAAATTAATACT 834
QY      286 -----IleGluGluLeuArgLysHisAlaLysIle 295
      835 GATGCATACAAAGAAATTTATTAAGAGTAAAGTA 870
Db

RESULT 5
US-09-252-991A-13635
; Sequence 13635, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

```

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13635
; LENGTH: 2097
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13635

Alignment Scores:
Pred. No.: 7,456-17 Length: 2097
Score: 211.50 Matches: 77
Percent Similarity: 46.18% Conservative: 56
Best Local Similarity: 26.74% Mismatches: 100
Query Match: 13.91% Gaps: 55
DB: 4 Gaps: 14

US-10-039-183a-2 (1-299) x US-09-252-991A-13635 (1-2097)
QY      48 ValAspGlyArgProIleThrLysSerAspPheAsp---MetIleLysGlnArgAsnPro 66
      610 GTGACGGCAAG---TTCAACGGGATGCTTCGACACAGTATCGCCAGATGAC--- 663
Db      67 AsnPheAspPheAspLysLeuLysGlu-----LysGluAlaLeuIleAsp 83
      664 -----TACAGCCGATGACATGCTCCGACAGATGCTCGCCAGAAATGCTATCGGC 714
QY      84 GlnAlaIleArgThrAlaLeu-----ValGluAsnGlu----- 94
      715 CAG---CTTCGGCGCGCTGCGCGGCAACCGTTTCGACCGACAAAGAAATTCGAGTCC 771
QY      95 ---AlaLysThrGluLysLeuAsp----- 101
      772 TTCGCTCGCTCGAAGAGCAGACCCGCCAATTCGCCACCTCGGCATCAAGCCGACGCT 831
Db      102 -----SerThrProGluPheLysAlaMetMetGluAlaValLys----- 114
      832 TCCAAGACGAGCGTACGAGAGAGAGGAAAGGCTTCTACGAAGGCCACAAAGAGCGAG 891
QY      115 -----LysGlnAlaLeuValGluPheThrAlaLysLysGlnAlaGluGluVal 130
      892 TTCATGACTCCCGACAGAGTGGTCTCGAATTCGTGAACCTGAAGAAAGTCTCTTCTTC 951
QY      131 LysLysValGlnIleProGluLysGluMetGlnAspPheTyraAsnAlaAsnLysAspGln 150
      952 GACCAAGTCACAGTACAGAGAGAGAGACCTCGAGCGCTGTAC-----CAGAAAGAAATC 1005
Db      151 LeuPheValLysGlnGluAlaHisAlaArgHisIleLeuValLysThrGluAsp----- 168
      1006 GCCAACCTCTCCGACAGAGCGGATGCCGCCACATCTGATGAGGTGAACAGACAAAGTCC 1065
QY      169 -----GluAlaLysArgIleIleSerGluIleAspLysGlnProLysAlaLysLys 185
      1066 GCGCAGCAGCAGCGCCAAAGCGGAATCGACGAGATC-----AAGCTCGCTCGCCAG 1119
Db      186 GluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAla 205
      1120 GCGCAGATTTGCGCGCGCTGCGCAAGAGATTTCCACAGATATGCGCTG-----GCC 1173
QY      206 GlnAsnGlyLysPheLysGlyLysPheGlnLysAsnGlnMetAlaProAspPheSerLys 225
      1174 GCCACCGCGCGGACCTGGGCTACCGCGCGCGCGGTACGACCCCGCTTCAGAGAG 1233
QY      226 AlaAlaPheAlaLeuThrProGlyAspTyThrLysThrProValLysThrGluPheGly 245
      1234 GCGCGTATGCGCTGACCAAGGTGAG--GTATCCGCGCGCGGTGAACATCCGTCGCGC 1290
QY      246 TyrHisIleIleTyLeuIleSerLysAspSerProValThrTyThrTyGluGlnAla 265
      1291 TACCACTGATCAAGCTCTGGCGCTGACGAGGCGCGGAAGTACGAGCCTGGAAGCCTC 1350

```

QY 266 LysProthrlleLysGlyMetLeuGlnGluLysLeuPheGlnGluArgMetAsnGlnArg 285
DB 1351 AGCCGAAGCTCGAGAGCAACTGAGAAACAGATGTCGACGACGCTTCCTCGAGGCT 1410
QY 286 lIegluGluLeuArgLysHisAla 293
DB 1411 ACCAAGACTTGGAAGCTCCGCC 1434

RESULT 6
US-09-252-991A-13809/C
Sequence 13809, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13809
LENGTH: 2331
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13809

Alignment Scores:
Pred. No.: 8,786-17 Length: 2331
Score: 211.50 Matches: 77
Percent Similarity: 46.18% Conservative: 56
Best Local Similarity: 26.74% Mismatches: 100
Query Match: 13.91% Indels: 55
Gaps: 14

US-10-039-183a-2 (1-299) x US-09-252-991A-13809 (1-2331)

QY 48 ValAepGlyArgProIleThrLysSerAepPheAsp--MetIleLysGlnArgAsnPro 66
DB 1899 GTCGACGGCAAG---TTCAACGGCGATCGCTTCGACAGATCATCGCCAGATGAAC--- 1846
QY 67 AenPheAepPheAspLysLeuLysGluLysGlu-----LysGluAlaLeuIleAsp 83
DB 1845 -----TACAGCCGCATGCAATGCCAGATCTCGCGCAGAAATGCTCATTCGCGC 1795
QY 84 GlnAlaIleArgThrAlaLeu-----ValGluAsnGlu----- 94
DB 1794 CAG---CTTCGCGCGCGCTTCGCGCGCGACCGCTTCGTCACCGACAGAAATTCAGTCC 1738
QY 95 --AlaLysThrGluLysLeuAsp----- 101
DB 1737 TTCGCTCGCTCGAGAGAGACGACCGGCACTTCGCCACCTGGCGATCAAGGCCGACGCT 1678
QY 102 -----SerThrProGluPheLysAlaMetMetGluAlaValLys----- 114
DB 1677 TCCAAAGACGAGCGTGCAGAGAGAGCGAGGCTTCCTACGAAAGGCCCAAGAGCGAG 1618
QY 115 -----LysGlnAlaLeuValGluPheTrpAlaLysGlnAlaGluGluVal 130
DB 1617 TTCATGACTCCCGAGAGAGGTGCTGTCGATACGTGGAACCTGAGAAATCTCTTCCTTC 1558
QY 131 LysAlaValAlaIleProGluLysGluMetGlnAspPheTrpAsnAlaAsnLysAspGln 150
DB 1557 GACCAAGGTGAGGTGAAGAGAGACCTCGAGGCGCTTAC-----CAGAAGAAATC 1504
QY 151 LeuPheValLysGlnGluAlaHisAlaArgHisIleLeuValLysThrGluAsp----- 168
DB 1503 GCCAACTCTCCGACGAGCGAGTCCGCCCATCTCTAGAGGTGAAGCAAGAGTCC 1444
QY 169 -----GluAlaLysArgIleIleSerGluIleAspLysGlnProLysAlaLysLys 185

DB 1443 GCGCAGCAGACGAGCCAGGCGGAATCGACGATC-----AAGCTCGCTCGGCCAAG 1390
QY 186 GluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAla 205
DB 1389 GCGCAGGATTCGCGCGCTGCGCCAAAGAGTTCTCCAGGATATCGGCTCG-----GCC 1336
QY 206 GlnAsnGlyLysPheGluLysPheGlnLysPheGlnMetAlaProAspPheSerLys 225
DB 1335 GCCACCGGCGGAGCCTGAGCTTACGCCCGCGCGCTGATACGACCCCGCTTCAGAGAG 1276
QY 226 AlaAlaPheAlaLeuThrProGluLysPheTrpLysThrProValLysThrGluPheGly 245
DB 1275 GCGCTGATGCGCTCAAGAGAGGTAG---GATCGCGCCCGGTGAAGCTCGTACGCGC 1219
QY 246 TyrHisIleIleLysLeuLysSerLysAspSerProValThrTrpLysThrGluGlnAla 265
DB 1218 TACCACTGATCACTGCTGCGCGCTGCGAGCGCGGAAATACGAGCCTGGAAGCCCTC 1159
QY 266 LysProthrlleLysGlyMetLeuGlnGluLysLeuPheGlnGluArgMetAsnGlnArg 285
DB 1158 AGCCGAAGCTCGAGAGCAACTGAGAAACAGATGTCGACGACGCTTCCTCGAGGCT 1099
QY 286 lIegluGluLeuArgLysHisAla 293
DB 1098 ACCAAGACTTGGAAGCTCCGCC 1075

RESULT 7
US-09-596-002-33
Sequence 33, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Patterson, Chandra
APPLICANT: Berg, Kim, L.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
FILE REFERENCE: PM-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
CURRENT FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 33
LENGTH: 63563
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte template ID No. 6632636 33
PUBLICATION INFORMATION:
US-09-596-002-33

Alignment Scores:
Pred. No.: 1,496-14 Length: 63563
Score: 211.50 Matches: 68
Percent Similarity: 45.07% Conservative: 60
Best Local Similarity: 23.94% Mismatches: 111
Query Match: 13.91% Indels: 45
Gaps: 11

US-10-039-183a-2 (1-299) x US-09-596-002-33 (1-63563)

QY 50 GlyArgProIleThrLysSerAepPheAspMetIleLysGlnArg-----AsnProAsn 67
DB 11755 GGCAGACCT-----TCATCTGATGCTCAATTATCAAGCAAGCATTAAGTAAGAAAT 11808
QY 68 PheAspPheAspLysLeuLysGluLysGluLysGluAlaLeuIleAspGlnAlaIleArg 87
DB 11809 TTTGACAGCACCATCTATCGCTCAAGCCACCAACTTTTACAAATATTCATCTGTAAT 11868
QY 88 ThrAlaLeuValGluAsnGluAlaLysThrGluLysLeuAspSerThrProGlu----- 105


```

QY 167 -----GluApGluAlaLysArgIleIleSer 175
DB 4816 AAATCTAAGAAAAGGACMAAGAGCTTAGATGATTAAGAACGCAACAAAGCTGA 4875
QY 176 GluIleApGlnProLysAlaLysGluAlaLysPheIleGluLeuAlaSerArg 195
DB 4876 GAAATTCACAAAAGAA---GTTTCAAAAGATCCAGTAATTTGGTGAATTCCTAAAAAA 4932
QY 196 AspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyLysAsnGlyLysPheGln 215
DB 4933 GAATCAATGATGATCTGCTTCA-----GCTAAAAAGATGCGAATTAAGTTATGTTCTT 4986
QY 216 LysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTyr 235
DB 4987 AAAGACAAACTGATTAAGATTTTGAAGAAAGCATTTAACCTTAAGATGCTGAAGTA 5046
QY 236 ThrLysThrProValLysThrGluPheGlyTyrHisIleIleTyrLeuIleSerLysAsp 255
DB 5047 TCAGAGGTT---GTTAAATCAAGCTTGGATATCATATT-----ATTAAAGCTGAT 5094
QY 256 SerProValThrTyrThrTyrGluGlnAlaLysProThrIleLysGlyMetLeuGlnGlu 275
DB 5095 AAACCAACAGACTTTAAACAGTGAA-----AAACAAAGCCTGAAGAA 5136
QY 276 LysLeuPheGlnGluArgMetAsnGlnArgIleGluGluLeuArg-LysHisAlaLysIle 295
DB 5137 AAATTAGTCGATCAGAAAGTACAAATAATTTTACGATGATGATCAAAAGT 5196
QY 295 e 295
DB 5197 C 5197

RESULT 9
US-08-781-986A-168
; Sequence 168, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunach
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 168:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7963 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-168

```

```

Alignment Scores:
Pred. No.: 2,62e-15 Length: 7963
Score: 206.50 Matches: 84
Percent Similarity: 43.30% Conservative: 55
Best Local Similarity: 26.17% Mismatch: 108
Query Match: 13.59% Indels: 74
DB: 4 Gaps: 15

US-10-039-183a-2 (1-299) x US-08-781-986A-168 (1-7963)
QY 5 IleLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMetAlaLysProAla 24
DB 4369 ATGATTAACAATTAATTAATGTTCCGGTACAGCTAGCTTATTAATTAGCCCTGTGGC 4428
QY 25 HisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrThrAspSerSerAlaVal 44
DB 4429 GCTAGTGCACAGACTCTTAAAGAAATTAACA-----TTAATTTCTTAAAGCTGAGAC 4482
QY 45 LeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMetIleLysGlnArg 64
DB 4483 GTAACAGTTGCAGAT-----ACAATGAAAAAATCGTAAAGATCAAAATT-----GCA 4530
QY 65 AsnProAsnPheAsp-----PheAspLysLeuLysGluLysGlu 77
DB 4531 AATGATCATTTTACTGAAATGTTAAATTAATTTTACGATGATTAATTAATTAATTAAGTT 4590
QY 78 LysGluAlaLeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGluAlaLysThr 97
DB 4591 AATGATTAAGAAAGATTGACCAAAATTT-----GAAAAAATGCA 4629
QY 98 GluLysLeuAspSerThrProGluPheLysAlaMetMetGluAlaValLysGlnAla 117
DB 4630 AAGCAATACGGCGGTAAAGATTAATTTGAA-----AAGGCCCTTCAACAGCAGGT 4680
QY 118 LeuValGluPheTrpAlaLysLysGlnAlaGlnGluValLys----- 131
DB 4681 TTAACA-----GCCGATTAATTAATTAAGAAATTTACGATGCTGCTTATCATTA 4731
QY 132 -----LysValGlnIleProGluLysGluMetGlnAspPheTyrAsnAla 146
DB 4732 GAATTAATCAATGATTAATTAATTAATCTCTGCTGAAATTAAGAA----- 4779
QY 147 AsnLysAspGlnLeuPheValLysGlnGluAlaHisAlaArgHisIleLeuValLysThr 166
DB 4780 -----GACAGCARGAAGCTTACACATTTTAATTAAGTT 4815
QY 167 -----GluAspGluAlaLysArgIleIleSer 175
DB 4816 AAATCTAAGAAAAGGACMAAGAGCTTAGATGATTAAGAACGCAACAAAGCTGA 4875
QY 176 GluIleApGlnProLysAlaLysGluAlaLysPheIleGluLeuAlaSerArg 195
DB 4876 GAAATTCACAAAAGAA---GTTTCAAAAGATCCAGTAATTTGGTGAATTCCTAAAAAA 4932
QY 196 AspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyLysAsnGlyLysPheGln 215
DB 4933 GAATCAATGATGATCTGCTTCA-----GCTAAAAAGATGCGAATTAAGTTATGTTCTT 4986
QY 216 LysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTyr 235
DB 4987 AAAGACAAACTGATTAAGATTTTGAAGAAAGCATTTAACCTTAAGATGCTGAAGTA 5046
QY 236 ThrLysThrProValLysThrGluPheGlyTyrHisIleIleTyrLeuIleSerLysAsp 255
DB 5047 TCAGAGGTT---GTTAAATCAAGCTTGGATATCATATT-----ATTAAAGCTGAT 5094
QY 256 SerProValThrTyrThrTyrGluGlnAlaLysProThrIleLysGlyMetLeuGlnGlu 275
DB 5095 AAACCAACAGACTTTAAACAGTGAA-----AAACAAAGCCTGAAGAA 5136
QY 276 LysLeuPheGlnGluArgMetAsnGlnArgIleGluGluLeuArg-LysHisAlaLysIle 295

```

```

Db      5137 AAATTAGTCGATCAGGAAGTACAAATAATCCAAATATTATTGACTGATGATACAAAGAT 5196
Qy      295 e 295
Db      5197 C 5197

RESULT 10
US-09-540-236-1818
; Sequence 1818, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NOCILEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709-2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1818
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: M. catarrhalis
US-09-540-236-1818

Alignment Scores:
Pred. No.: 1,65e-16 Length: 1110
Score: 205.50 Matches: 67
Percent Similarity: 44.01% Conservative: 58
Best Local Similarity: 23.59% Mismatches: 114
Query Match: 13.52% Indels: 45
DB: 4 Gaps: 10

US-10-039-183A-2 (1-299) x US-09-540-236-1818 (1-1110)
Qy      50 GlyArgProlIethrInLysSerAspPheAspMetIleLysGlnArg-----AsnProAsn 67
Db      238 GGCAGACT-----TCATCTGATGCTCAATTTATCAAGCAAGCATTTAGATTAAGAAAT 291
Qy      68 PheAspPheAspLysLeuLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 87
Db      292 TTTCAGACGACCAATCTATATCGCTTCAGACGACCAACCTTACCAAAATATCATGTAAT 351
Qy      88 ThrIleuValIleGluSerGlnAlaLysThrGlnLysLeuAspSerThrProGlu----- 105
Db      352 GGCCTAGCAATAGACCCCAAGACGATGCTCCGAAAGCGATATCATCTCGTAATAGC 411
Qy      106 -----PheLysAlaMetMetGlnLysValLysGlnAlaValLysGlnAlaValGlu 120
Db      412 CAAGAAAGACATTTATTTTTCAGAGCTCAAGCCTTAGTATTGCTGAATTTCTGCTCAA 471
Qy      121 -----PheTrpAlaLysLysGlnAlaGluLysValLys 131
Db      472 GAAGTATTGGGTGCTTCGGGAGAGGGCGTGGCTAGAGATTAAGAAAGTATACATCGCT 531
Qy      132 LysVal-----GlnIleProGluLys--GluMetGlnAspPhe 143
Db      532 AAGCTGTAGACAGCAAGTCATCGCCAAATCTCCAGACAGGAGGAGCTGTCCAGAAATTAC 591
Qy      144 TyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisIleArgHisIleLeu 163
Db      592 TATCATGTCTAATCAAGTTGAAATTGTTTATGTTATGCCAATATAGACGCTTCGCCAATTTG 651
Qy      164 -----ValLysThrGlnAspGlnLysValLysArg 172
Db      652 CTTCGGTGTCCACCAACAGAGGAGGTAAGAGCGCATTTGAATTAATAAAGACGCTGGCAA 711
Qy      173 IleIleSerGlnIleAspLysGlnProLysAlaLysLysGlnLysPheIleGlnLeu 192
Db      712 TTGATTCATCGCTTAAATCAA-----AGCCAAATCCAGATAGTATTTATTTAGAGTTT 765
Qy      193 AlaAsnArgAspTrnIleAspProAsnSerLysAsnIleGlnAsnGlyLysAspLeuGly 212
Db      766 GCATGTGATATCTGCTGCTGCTGCTTCAAAA-----GATGATGGCGGTGAACCTGGGT 816

```

```

OY      213.  LysPhegNlysaNgImMetalAPRoaspPheSerLySaIaAlaPhealAleuNhPro 232
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      817  GTATTACAAAAGGCGATACCGGTGCTTAAATTGAGAGTCCGTAATTGCTTGG---CCA 873
OY      223  GLyAsPryThrLyThrProValyThrGluPheGlyTyRhiAlleIeTyRleuIle 252
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      874  GTTGGTATCAGTATTAAATCCATTCGATTGAACGGGATATGGGATCCATGTAATGAAGTTCA 933
OY      253  SerLySaPserProValThrTyThrTyRgluGlnAlaIysProThrIlelyseGlyMet 272
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      934  CAAAAGCGAGGAGGTGAGCGACGTGCTTTGAAGGCGCATATCCATATCTGAATAATCAT 993
OY      273  LeuGlnGluLySleuPheGlnGluArgMetAsnGlnArgIleGluGluLeuArgLySHis 292
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      994  CTAAAGCAGCGATCATTTTCATCATATGTTGTGTACTATTATTTTGAATTTGAGCCAAAAG 105
OY      293  AlaLySleVal 296
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1054  GCTGATATTATT 1065

RESULT 11
US-09-902-540-3927
; Sequence 3927, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wisgand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(1,5849)B
; CURRENT FILING DATE: US/09/902,540
; PRIOR APPLICATION NUMBER: 2001-07-10
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3927
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3927

Alignment Scores:
Pred. No.:      1,366-15      Length:      1,371
Score:          199.50      Matches:      77
Percent Similarity: 44.41%      Conservative: 54
Best Local Similarity: 26.10%      Mismatches: 95
Query Match:      13.12%      Indels:      69
DB:              4      Gaps:      14

US-10-039-183A-2 (1-299) x US-09-902-540-3927 (1-1371)

OY      53  lIethrLySserAPhAspMecIleLyS-----GlnArg 64
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      130  ATCCGCCGCTCCGACGATGAGCTGCGCAAGCTCATCCACAGAACCGGACTTCCAGAG 189
OY      65  AsnProAsnPhAspPhAspLySleuLySgu----- 75
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      190  GATGGCCAGTTCGATTCGAGGCTACCAAGCAGGCTGCGGACTTTCACCGGAAGTCC 249
OY      76  ---LysGluLySglnAlaLeuIleAspGlnAlaIleArgTyRAlaLeuValGluAsnGlu 94
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      250  CCGAGGACCTTCGAGGCGGAGCTGCGCGGCAATATGGCGGCCAGAAAGATGATGACGCG 309
OY      95  AlaLySthrGluLySleuAspSerThrProGluPheLySAlaMetMetGluAlaValIys 114
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      310  GTGGCGGCGCAACGCGGTGTGTGTGCGACGACGAGGTCCGCGCCGCTTCGAGAAAGAGGC 369
OY      115  LysGlnAlaLeuValGluPhe-----TrpAlaLySlys----- 125
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      370  AACCAAGCGAAGGTCTCTTTCGCGCGCTTCTCGCGGCGGCAATGTACGCGCAAGGTCCCC 429

```



```

CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-557-884-1
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
Alignment Scores:
Pred. No.: 2.77e-08 Length: 1830121
Score: 180.50 Matches: 81
Percent Similarity: 42.90% Conservative: 64
Best Local Similarity: 23.96% Mismatches: 129
Query Match: 11.88% Indels: 64
DB: 4 Gaps: 16
US-10-039-183A-2 (1-299) x US-09-557-884-1 (1-1830121)
QY 1 MetLysLysAsnIleLeuLeuValGlyAlaLeuSerThrSerPheLeu--- 19
DB 1067561 CTTGCTCAAAACATGCTTAATGATGATTCAGAAATTAATTCTCGCAATATGTTAA 1067620
QY 20 -----MetAlaLysProAla----- 24
DB 1067621 GAATTAAATTAGGCGTGAGTGAATGAATGAATTAACGTCGATCGCATCTTAT 1067680
QY 25 -----HisAlaAlaAsnAlaThrHis-----AsnThrLys 35
DB 1067681 TTTCAAGTAAGGTAATTTGATTAATGCTTTTATCAAGATATTATCAACAAATCAT 1067740
QY 36 LysThrThrAspSerSerLacIlyValLeuAlaThr-----Val 48
DB 1067741 TTAACTTCAGATGCTTACCGCTTCAATTTTACGCTTCTGTTACCTTGAACAAATACAA 1067800
QY 49 AspGlyArgProIleThrLysSerAspPhe-----AspMetIleLysGlnArgAsn 65
DB 1067801 AATGCG-----GTTGCTACAGGTGAATTTATTTCTTCGCTCAAGTAAATAATAGCGGA 1067854
QY 66 ProAsnPheAspPheAspLysLeuLysGlnLysGlnLysGlnAlaLeuIleAspGlnAla 85
DB 1067855 GAAGTTTCTTCAAAAGCGTTAGCTGTTAGCAACTTATCTTCCATGCAATGAATG 1067914
QY 86 IleArgThrAlaLeuValGlnGlnAlaLysThrGlnLysLeuAspSerThrProGlu 105
DB 1067915 GCATAACATACGTCTCTGATGATGAATCAAAAGCTATTATGAAGCTAATCAAAAGTCT 1067974
```

```

QY 106 PheLysAlaMetMetGlnAlaValLysLysGlnAlaLeuValGlnPheTrpAlaLysLys 125
DB 1067975 TTT---GTTCAAGCCAGACAGGTTAAAGTTCAG-----TATATTGATCTT 1068016
QY 126 GlnAlaGlnGluVal---LysLysValGlnIleProGluLysGlnMetGlnAspPheTr 144
DB 1068017 TCTGCAGATAACATCATGTAAGAAATCTTCAAGTTACAGATGTAAGAAATTCGCAATATAT 1068076
QY 145 AsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArgHisIleLeuVal 164
DB 1068077 CAAGTATATTAAGCACAA---TTTATGACTCAA-----CATTTAGCTCATATTCAATT 1068127
QY 165 LysThrGluAspGlnAlaLysArgIleIleSerGlnIleAspLysGlnProLysAlaLys 184
DB 1068128 GCAAAATGAACAAGATGCGAAAGTGCTTACAGAGAAATTA-----CAG 1068169
QY 185 LysGlnAlaLysPheIleGlnLeuAlaAsnArgAspThrIleAspProAsnSerLysAsn 204
DB 1068170 AAAGGGGCAAAATTTGCTGATGATGCTTAAGCGAAATCTTTGGATTAATAATTTCT--- 1068223
QY 205 AlaGlnAsnGlyGlyAspLeuGlyLysPheGlnLysAsnGlnMetAlaProAspPheSer 224
DB 1068224 GGTGGAATGCGCGGGGATTTAGCTGGGTAATATGGAATGAATTACCAAAAGCTTTGAA 1068283
QY 225 LysAlaAlaPheAlaLeuThrProGlyAspThrLysThrProValLysThrGlnPhe 244
DB 1068284 GATGCCGACGCTGCTTACAGGTAGGCAATATAGCCAA---CCAAATCAATGTGATGCGC 1068340
QY 245 GlyThrHisIleIleLysLeuLysSerLysAspSerProValThrThrThrGlnGln 264
DB 1068341 AATTACCAATATGTTGTTAGTCAAGAACGTAAAGCG-----CAAAAGTTTGAGAAAT 1068391
QY 265 AlaLysProThrIleLysGlyMetLeuGlnGlnLysLysLeuPheGlnLysArgMet----- 282
DB 1068392 GTGAAAGCTCAAAATTCAGATTTAGTCTGTAATCTTTGATGAGAGAGTCGTTATTTCT 1068451
QY 283 -----AsnGlnArgIleGlnGlnLeuArgLysHisAlaLysIleValIle 297
DB 1068452 TTAGAAAAACAAGCAGATGATTAAGGCTTTTGAAGACAGCAAGTCTTAAATA 1068503
RESULT 14
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
```

```

1 APPLICATION NUMBER: 08/426,787
2 FILING DATE: 1995-04-21
3
4 ATTORNEY/AGENT INFORMATION:
5 NAME: Kenley K. Hoover
6 REGISTRATION NUMBER: 40,302
7 REFERENCE/DOCKET NUMBER: PB186P1C1
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 301-610-5790
10 TELEFAX: 310-309-8439
11
12 INFORMATION FOR SEQ ID NO: 1:
13
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 1830121 base pairs
16 TYPE: nucleic acid
17 STRANDEDNESS: double
18 TOPOLOGY: linear
19
20 SEQUENCE DESCRIPTION: SEQ ID NO: 1:
21
22 US-09-643-990A-1
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005

```

[illegible]

```

Db      325 GCCAATCCGGGGCGCTTGCTCGGCCCTTCGACCTGCAATGCGCCATGCTGCTGCC 384
Qy      165 -----LysThrGluAspGluAlaLysArgIleIle 174
Db      385 TCCGGCGCGGACGACCTGGAAGCCGCCGACGCGCGCAAGACGCGCGAGCTGCTC 444
Qy      175 SerGluIleAspLysGlnProLysAlaLysGluAlaLysPheIleGluLeuAlaAsn 194
Db      445 GACGAGTTGCGCGGATCCCGAG-----CGCTTCGTGACCTGGGCCCGG 489
Qy      195 ArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyGlyAspLeuGlyLysPhe 214
Db      490 CGTTTCTCGCCTGCTCG---TCGAAGAGGCG-----GGCGCGACCTCGGCTGGAATC 540
Qy      215 GlnLysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAsp 234
Db      541 GAGCCGGGGCAGACCGTCCCGAATTCGAGAGCGCTGCTGCCGCCGCCGCCGCGG--- 597
Qy      235 TyrThrLysThrProValLysThrGluPheGlyTyrHisIleIleTyrLeuIleSerLys 254
Db      598 CTGCTCGAGCATCCGCTGGAAGCCGCTATGCGCTGCTGATGTGTGAACTGTGGGCGCGC 657
Qy      255 AspSerProValThrTyrThrTyrGluGlnAlaLysProThrIleLysGlyMetLeuGln 274
Db      658 GAAGCGCGGACCGCTCGACTTCGACGCGCGCGCACAGATCGCCCGCACCTGCGAG 717
Qy      275 GlnLysLeuPheGlnGluArgMetAsnGlnArgIleGluGluLeuArgLysHisAlaLys 294
Db      718 GTGCAAGTGTTCAGAGCGAGCGGTGGACAGTACATCGCGCTGCGCGGAGACGCTGT 777
Qy      295 Ile 295
Db      778 ATC 780

```

Search completed: April 24, 2005, 11:43:22
 Job time : 1174.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2005, 05:52:31 ; Search time 583.007 Seconds

(without alignments)
3113.588 Million cell updates/sec

Title: US-10-039-183a-2

Perfect score: 1520

Sequence: 1 MKNINLALVGLSTSLFM.....ERMQRIBELRKHATVINK 299

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO/US10039183/runat_22042005_122426_20167/app_query.fasta_1.1038
-DB=Published Applications NA -QMT=faefap -SUFFIX=trpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=spct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USR=US10039183 @CGN_1.1.894 @runat_22042005_122426_20167
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1520	100.0	1000	9 US-09-881-752A-243	Sequence 243, App
2	1520	100.0	1149	8 US-08-831-310-1	Sequence 1, Appl
3	1520	100.0	1149	15 US-10-039-183A-1	Sequence 1, Appl
4	1486	97.8	900	17 US-10-335-977-483	Sequence 483, App
5	745.5	49.0	504	17 US-10-335-977-482	Sequence 482, App
6	679	44.7	456	17 US-10-335-977-481	Sequence 481, App
7	221.5	14.6	1720	17 US-10-398-221-3354	Sequence 3354, App
8	221.5	14.6	43980	17 US-10-398-221-5	Sequence 5, Appl
9	221.5	14.6	3011208	17 US-10-398-221-2058	Sequence 2058, Ap
10	211.5	13.9	63563	17 US-10-672-787-33	Sequence 168, App
11	206.5	13.6	7963	8 US-08-781-986A-168	Sequence 168, App
12	206.5	13.6	7963	17 US-10-329-624-168	Sequence 168, App
13	205.5	13.5	960	19 US-10-470-048B-173	Sequence 173, App
14	200.5	13.2	1163	17 US-10-398-221-1250	Sequence 1250, Ap
15	191.5	12.6	495269	17 US-10-398-221-8	Sequence 8, Appl
16	191.5	12.6	3011208	17 US-10-398-221-2058	Sequence 2058, Ap
17	182.5	12.0	6157	17 US-10-398-221-3752	Sequence 3752, Ap
18	180.5	11.9	1830121	17 US-10-329-670-1	Sequence 1, Appl
19	180.5	11.9	1830121	18 US-10-158-865-1	Sequence 1, Appl
20	162.5	10.7	486	9 US-09-974-300-1374	Sequence 1974, Ap
21	162	10.7	638	9 US-09-974-300-6328	Sequence 6328, Ap
22	159.5	10.5	403	19 US-10-899-942-47	Sequence 47, Appl
23	156.5	10.3	942	18 US-10-398-186-15	Sequence 15, Appl
24	156.5	10.3	1830121	17 US-10-329-670-1	Sequence 1, Appl
25	156.5	10.3	1830121	18 US-10-158-865-1	Sequence 1, Appl
26	156.5	10.3	2731748	18 US-10-297-465A-1	Sequence 1, Appl
27	155.5	10.2	864	17 US-10-275-026A-53	Sequence 53, Appl
28	155.5	10.2	864	17 US-10-275-026A-171	Sequence 171, App
29	142.5	9.4	279	13 US-10-027-806-67	Sequence 67, Appl
30	142.5	9.4	279	13 US-10-027-806-67	Sequence 67, Appl
31	142.5	9.4	279	14 US-10-027-801-67	Sequence 67, Appl
32	142.5	9.4	279	16 US-10-029-120-67	Sequence 67, Appl
33	142.5	9.4	32998	13 US-10-027-806-1	Sequence 1, Appl
34	142.5	9.4	32998	13 US-10-034-623-1	Sequence 1, Appl
35	142.5	9.4	32998	14 US-10-027-801-1	Sequence 1, Appl
36	142.5	9.4	32998	16 US-10-029-120-1	Sequence 1, Appl
37	139.5	9.2	648	19 US-10-899-942-55	Sequence 55, Appl
38	134.5	8.8	42432	13 US-10-027-806-2	Sequence 2, Appl
39	134.5	8.8	42432	13 US-10-034-623-2	Sequence 2, Appl
40	134.5	8.8	42432	14 US-10-027-801-2	Sequence 2, Appl
41	134.5	8.8	42432	16 US-10-029-120-2	Sequence 2, Appl
42	133.5	8.8	707	19 US-10-899-942-58	Sequence 58, Appl
43	131.5	8.7	584	18 US-10-021-323-15277	Sequence 15277, A
44	131.5	8.7	1242	17 US-10-335-977-4082	Sequence 4082, Ap
45	130.5	8.6	484	10 US-09-918-995-22268	Sequence 22268, A

ALIGNMENTS

RESULT 1
US-09-881-752A-243
Sequence 243, Application US/09881752A
Patent No. US20020115078A1
GENERAL INFORMATION:
APPLICANT: Kleantous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
FILE REFERENCE: 06132/041002
CURRENT FILING DATE: 2001-06-15
PRIOR FILING DATE: 1997-04-01
NUMBER OF SEQ ID NOS: 370
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 243
LENGTH: 1000
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (51)...(947)
US-09-881-752A-243

Alignment Scores:

Pred. No.:	4,93e-137	Length:	1000
Score:	1520.00	Matches:	299
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-039-183A-2 (1-299) x US-09-881-752A-243 (1-1000)

QY 1 MetLysLysAsnLLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 51 ATGAAAAAATAATCTTAATTTAGCGTTAGCGGCGCTTGAGCAGCGCTTTTGATG 110
QY 21 AlaLysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrLysPse 40
DB 111 GCTAAGCCGGCTCATTAACGCAATTAACGCTAACACGAAAAAAGCACTGATTCT 170
QY 41 SerAlaGlyValLeuAlaThrValAspGlyArgProLLeThrLysSerAspPheAspMet 60
DB 171 TCAGACGGGCTGTAGCGACAGTGGATGGACAGCCATACCTAAAGCGATTGTGACATG 230
QY 61 IleLysGlnArgAsnProAsnPheAspPheAspLysLeuLysGlnLysGlnLysGln 80
DB 231 ATTAAGCAACGAAATCTTAATTTGATTTTGACAGCTTAAAGAGAAAAAAGAGAGCC 290
QY 81 LeuLLeaPGLaLleLLeaGThraLLeuValGluAsnGlnAlaLysThrGluLysLeu 100
DB 291 TTGATTTGATCAAGCTATTTCGACCCGCTTGTAGAAAAAGAAAGCTAAACCGAGAAATTG 350
QY 101 AspSerThrProGluPheLysAlaMetMetGluAlaValLysLysGlnAlaLeuValGlu 120
DB 351 GACAGCACCTCCAGAAATTTAAACGAGTGAAGCGGCTTAAAGAGAGAGAGAGAGAGAG 410
QY 121 PheTPrAlaLysLysGlnAlaGluGluValLysLysValGlnLLeProGluLysGluMet 140
DB 411 TTTTCGGCTTAAGAAACAGGCTGAAGAGTGAAGAAAAAGTCCAAATCCAGAAAAAAGAAATG 470
QY 141 GlnAspPheTyrAsnAlaAsnLysAspGlnLeuPheValLysGlnGlnAlaHisAlaArg 160
DB 471 CAAGATTTTTCACACGCTAACAAAGATCGCTTTTGTCAACAGCAAGAGCCCATGCTAGG 530
QY 161 HisLLeuValLysThrGluAspGluAlaLysArgLLeLLeSerGlnLLeAspLysGln 180
DB 531 CATATTTTAAGTAAACCGAAGATGAGGCTAAACGAGTAAATTTCTGAGATTGACAAACG 590
QY 181 ProLysAlaLysLysGlnAlaLysPheLLeGluLeuAlaAsnArgAspThrLLeAspPro 200
DB 591 CCAAGAGCTTAAGAAAGAGCTTAATTCATTTGAGTTAGCCAAACGGGATACGATTGATCT 650
QY 201 AsnSerLysAsnAlaGlnAsnGlyLysAspLeuGlyLysPheGlnLysAsnGlnMetAla 220
DB 651 AACACAGAGAAACGCGCAAAATGCGGTGATTTGGGGAATTTCCAAAGAACCAATGCGCT 710
QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTyrThrLysThrProVal 240
DB 711 CCGGATTTTTCATAAGCCGCTTTCGCTTAACTCTCGGGGATTTACATAAAACCCCTGTT 770
QY 241 LysThrGluPheGlyTyrHisLLeLLeLLeLLeLLeLLeLLeLLeLLeLLeLLeLLeLLe 260
DB 771 AAAACAGAGTTGGTTATCATATTATCTATTGATTTCTAAAGATAGCCCTGTAACCTAT 830
QY 261 ThrTyrGluGlnAlaLysProThrLLeLysGlyMetLeuGlnLysLysLeuPheGlnGlu 280

DB 831 ACTTATGAACAGGCTAAACCTAACCTTAAGGGATGTTTACAGAAAAAGCTTTTCCAAGAA 890
QY 281 ArgMetAsnGlnArgLLeGluLLeuArgLysHisAlaLysLLeValLLeAsnLys 299
DB 891 CGCATTAATCAAGCATTTGAGGAACTAAGAAAGCAGCTTAATAATTTGTTATCAACAAG 947

RESULT 2

US-08-831-310-1
Sequence 1, Application US/08831310
Publication No. US20020026035A1

GENERAL INFORMATION:

APPLICANT: Kleantous, Harold et al.
TITLE OF INVENTION: Helicobacter GHPD 1360 and
TITLE OF INVENTION: GHPD 750 Polypeptides and Corresponding Polynucleotides
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,310
FILING DATE: 01-APR-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,175
REFERENCE/DOCKET NUMBER: 06132/037001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1149 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA

FEATURE:

NAME/KEY: Coding Sequence
LOCATION: 106...1002
OTHER INFORMATION:
NAME/KEY: Signal Sequence
LOCATION: 106...166
OTHER INFORMATION:

US-08-831-310-1

Alignment Scores:

Pred. No.:	5,86e-137	Length:	1149
Score:	1520.00 <td>Matches:</td> <td>299</td>	Matches:	299
Percent Similarity:	100.00% <td>Conservative:</td> <td>0</td>	Conservative:	0
Best Local Similarity:	100.00% <td>Mismatches:</td> <td>0</td>	Mismatches:	0
Query Match:	100.00% <td>Indels:</td> <td>0</td>	Indels:	0
DB:	8	Gaps:	0

US-10-039-183A-2 (1-299) x US-08-831-310-1 (1-1149)

QY 1 MetLysLysAsnLLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 106 ATGAAAAAATAATCTTAATTTAGCGTTAGCGGCGCTTGAGCAGCGCTTTTGATG 165
QY 21 AlaLysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysLysThrLysPse 40
DB 166 GCTAAGCCGGCTCATTAACGCAATTAACGCTAACACGAAAAAAGCACTGATTCT 225

DB 886 ACTTATGAACAGGCTAAACCTACCTAAGGGATGTTACAGAAAGCTTTCCAGAA 945
 QY 281 ArgMetAsnGlnAxyTlGluGluLeuArgLysHsAlaLysTlleValTleAsnLys 299
 DB 946 CGCATGAATCAACGCAATGAGAACTAAGAAAGCAGCTAAATTGTATCAACAAG 1002

RESULT 4

US-10-335-977-483

Sequence 483, Application US/10335977

Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

RELATING TO HELICOBACTER PYLORI FOR

DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 483:

SEQUENCE CHARACTERISTICS:

LENGTH: 900 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...900

SEQUENCE DESCRIPTION: SEQ ID NO: 483:

US-10-335-977-483

Alignment Scores:

Pct. No.: 8.21e-134 Length: 900

Score: 1486.00 Matches: 290

Percent Similarity: 99.00% Conservative: 6

Best Local Similarity: 96.99% Mismatches: 3

Query Match: 97.76% Indels: 0

DB: 17 Gaps: 0

US-10-039-183A-2 (1-299) x US-10-335-977-483 (1-900)

QY 1 MetLysLysAsnLleLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
 DB 1 ATGAAATTAATATCTTAATTAGCGTTAGGCGCGCTTGAATG 60

QY 1 MetLysLysAsnLleLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
 DB 1 ATGAAATTAATATCTTAATTAGCGTTAGGCGCGCTTGAATG 60

QY 21 AlaLysProAlaHsAsnAlaAsnAlaAlaThrHsAsnThrLysLysThrThrAspSer 40
 DB 61 GCTAAGCCGGCTCATTAATGCGAATTAATCTTAAGCTTAACAGAAAGAAACGATGCT 120

QY 41 SerAlaGlyValLeuAlaThrValAspGlyArgProIleThrLysSerAspPheAspMet 60
 DB 121 TCAGCAGGCGTGTGAGCAGAGTGATGAGAGCCATCACCAAAAGGATTTGATGATG 180

QY 61 IleLysGlnArgAsnProAsnPheAspPheAspLysLysGluLysGluLysGluAla 80
 DB 181 ATTAAGCAACGAATCTTAATTTGATTTGATTAAGCAAGCTTAAGAAAGAAAGAAAGACC 240

QY 81 LeuIleAspGlnAlaIleArgThrAlaLeuValGluAsnGluAlaLysThrGluLysLeu 100
 DB 241 TTGATTTGAGCAAGCTATCCGACCGCTTGAGAAATGAGGCTTACGAGCAAAAAGCTT 300

QY 101 AspSerThrProGluPheLysAlaMetThrGluAlaValLysLysGlnAlaLeuValGlu 120
 DB 301 AATCAGACTCCAGATTTTAAAGCGATGAGAGCGTTAAAGAAACAGGCTTTAGTGAA 360

QY 121 PheTrpAlaLysLysGlnAlaGluGluValLysLysValGlnIleProGluLysGluMet 140
 DB 361 TTTGGCTTAAGAAACAGGCTGAGAGATGAAAGATCCAAATCCAGAAAGAAATG 420

QY 141 GlnAspPheTyAsnAlaAsnLysAspGlnLeuPheValLysGlnGluAlaHsAlaArg 160
 DB 421 CAGGATTTTCAACGCAATTAAGATCAGCTTTTTCAGAGCAAGAACCCATCTAGG 480

QY 161 HisIleLeuValLysThrGluAspGluAlaLysArgIleIleSerGluIleAspLysGln 180
 DB 481 CATATTTTGAAGAAACCGAAGATGAGCCAAAGGATTAATTTGAGATGAGCAAGAG 540

QY 181 ProLysAlaLysLysGluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspPro 200
 DB 541 CCAAGGCTTAAGAAAGAAAGCCAAATTCATTAAGTTCAGCAATCCGGAATACATGCT 600

QY 201 AsnSerLysAsnAlaGlnAsnGlyGlyAspLeuGlyLysPheGlnLysAsnGlnMetAla 220
 DB 601 AACGCAAGAAACGGCAAAATGGCGGATTTGGGGAATTCAAAGAAACCAATGCT 660

QY 221 ProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTyThrLysThrProVal 240
 DB 661 CCGGATTTTCTTAAGCCGCTTTCCTTAACCTCGGGGATTTACACTTAAACCCCTGTT 720

QY 241 LysThrGluPheGlyTyHisIleIleTyIleuIleSerLysAspSerProValThrTyR 260
 DB 721 AAACAGAGTTGGTTATCATATTAATTTGATTTGATTTGATTAAGATAGCCCTGATCTTAT 780

QY 261 ThrTyGluGlnAlaLysProThrIleLysGlyMetLeuGlnLysLeuPheGlnGlu 280
 DB 781 ACTTATGAGCAAGCTTAACCTTAACCTTAAGGGATGTTACAAAGAAACCTTTCCAGAA 840

QY 281 ArgMetAsnGlnAxyTlGluGluLeuArgLysHsAlaLysTlleValTleAsnLys 299
 DB 841 CGCATGAATCAACGCAATGAGAAATTAAGAGCAGCTTAAATTGTATCAACAAG 897

RESULT 5

US-10-335-977-482

Sequence 482, Application US/10335977

Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

RELATING TO HELICOBACTER PYLORI FOR

DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 482:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...504
SEQUENCE DESCRIPTION: SEQ ID NO: 482:
US-10-335-977-482.

Alignment Scores:
Pred. No.: 1,046-62 Length: 504
Score: 745.50 Matches: 148
Percent Similarity: 91.76% Conservative: 8
Best Local Similarity: 87.06% Mismatches: 11
Query Match: 49.05% Indels: 3
Gaps: 3

US-10-039-183a-2 (1-299) x US-10-335-977-482 (1-504)

QY 131 LysylValylGlnIleProGluIleGluMetGlnAspPheTyrAsnAlaAsnLysAspGln 150
DB 1 AAAAAAGATCCAAATCCAGAAAAAAGAAATGACGATTTTTCACACGCCAATTAAGATCAG 60

QY 151 LeupheValylGlnGluAlaHisAlaArgHisIleLeuValIleLeuValIleThrGluAspGluAla 170
DB 61 CTTTGTGACGAAAGAAAGCCCATCTAGGCAATTTTGTAGAAAAACGAGATGAAGCC 120

QY 171 LysArgIleIleSerGluIleAspLysGlnProLysAlaLysLysGluAlaLysPheIle 190
DB 121 AAAAAAGATTATTTCTGAGATTGACAAACAGCCAAAGGCTTAAAAAGAAAGCCAAATTCATT 180

QY 191 GluLeuAlaAsnArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyGlyAsp 210
DB 181 GAGTTAGCCCAATCGGAGATACGATTGATCTTAACACAAAGAACGCCCAAAATGGCGGTGAT 240

QY 211 LeuGlyLysPheGlnLysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeu 230
DB 241 TTGGGAAATTCCTCAAAAGAACCAAAATGGCTCGGATTTTTCAAAAGCCGCTTTCCTTTA 300

QY 251 ThrProGlyAspTyrThrLysThrProValLysThrGluPheGlyTyrHisIleIleTyr 250
DB 301 ACTCTGGGAGATTACACTAAACCCCTGTAAAAACAGAGTTTGTTATCATATTATCAT 360

QY 261 LeuIleSerLysAspSerProValIleThrTyrThrTyrGluGlnAlaLysProThr---Ile 269
DB 361 TTGATTCTTAAAGATAGCCCTGTAACTTAATCTTATGAGCAAGCTTAAACCTACAGAGTTA 420

QY 270 LysGlyMetLeuGlnGluIleLysPheGlnIleArgMetLeuGlnArgIleGlnGluLeu 289
DB 421 AGGGAGATTATCCAGGAA---GCTTTCCAGACGATGAATGACGCGATG---GAATTA 474

QY 290 ArgLysHisAlaLysIleValIleAsnLys 299
DB 475 AGTAGCAGCTAAGTGTATCAACAAATAGA 504

RESULT 6
US-10-335-977-481
Sequence 481, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 481:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...456
SEQUENCE DESCRIPTION: SEQ ID NO: 481:
US-10-335-977-481

Alignment Scores:
Pred. No.: 2,376-56 Length: 456
Score: 679.00 Matches: 136
Percent Similarity: 96.60% Conservative: 6
Best Local Similarity: 92.52% Mismatches: 4
Query Match: 44.67% Indels: 1
Gaps: 0

US-10-039-183a-2 (1-299) x US-10-335-977-481 (1-456)

QY 1 MetLysLysAsnIleLeuAsnLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMet 20
DB 16 ATGAAAAAATATATCTTAATTTAGCGTGTAGCGGCGGCTTGAGACGCGCTGCTTTTGTATG 75

Db 21393 GCAGCTTATGCCCTTAAAAACAAGGCGACATCAGCGCTCCAGTAAAAAACAATATGGA 21

Db 2353051 GCTCAAGTGCTTAACTGAGAAATCATTC

QY 134 Gln-----IleProGluysGluMetGlnAspPheTyrAsn 145
 Db 2353111 CAAAAAGCACTGAAGCTAATGCTGANTACTAGCAGTAAAGCACTTAAAGAGTACTACAAA 2353170
 QY 146 AlaAsnIysAspArgGlnLeuPheValIysGlnGlnIuAaHisAlaArgHisIleLeuValIys 165
 Db 2353171 ACTTGCGAACCAAGATATTACTGTAAAGC-----CATATTCTAGTAGCT 2353212
 QY 166 ThrGluAspGluAlaIysArgIleIleSerGluIleAspIysGlnProIysAlaIysIys 185
 Db 2353213 GATGAAAACAAGCCAAAGAAAGATT-----GAACAAAAAAGCTTAAAGAC 2353254
 QY 186 GluAlaIysPheIleGlnLeuAlaAsnArgAspThrIleAspProAsnSerIysAsnAla 205
 Db 2353285 GCGCGAAATTTGCTGATTAGCAAAAGAAATATTCTACAGTACTGCTCAATAA----- 2353308
 QY 206 GluAsnGlyGlyAspLeuGlyIysPheGlnIysAsnGlnMetAlaProAspPheSerIys 225
 Db 2353309 GAAAAAGGTGGCCCAATTGACACCAATTGGTTGCTGTGTAATAAGATCTTCGATTGGAAAA 2353368
 QY 226 AlaAlaPheAlaLeuThrProGlyAspTyrThrIysThrProValIysThrGluPheGly 245
 Db 2353369 GCAGCTTATGCCCTTAAAAAAGAAAGGACATCAGCGCTCCAGTAAAGCAATATGGA 2353428
 QY 246 TyrHisIleIleIleTyrLeuIleSerIysAspSerProValThr--TyrThrTyrGluGln 264
 Db 2353429 TACCAACATCATCCCAATG-----GACAAACACGCTACAAAAACAACCTTTGATATA 2353479
 QY 265 AlaIysProThrIleIys---GlyMetLeuGlnIuIysLeuPheGlnIuArgMetAsn 283
 Db 2353480 GATAAAAAAGCTGTTAAAGAAAGCTTACTCTTCATCCCAATTACTACAGAAACATGCA 2353539
 QY 284 GlnArgIleGluGluLeuArgIysHisAlaIysIleValIle 297
 Db 2353540 AAAACGCTTAAAAAAGAAATCAACAAAGATGCTTACGTAAGTAAAGTT 2353581
 RESULT 10
 US-10-672-787-33
 ; Sequence 33, Application US/10672787
 ; Publication No. US20040067554A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LAGACE, Robert, E
 ; APPLICANT: PATTERSON, Chandra
 ; APPLICANT: BEGG, Kim, L.
 ; TITLE OF INVENTION: NOCLODITIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
 ; FILE REFERENCE: ELITRA.025C1
 ; CURRENT APPLICATION NUMBER: US/10/672,787
 ; CURRENT FILING DATE: 2003-09-26
 ; PRIOR APPLICATION NUMBER: 09/596,002
 ; PRIOR FILING DATE: 2000-06-16
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 33
 ; LENGTH: 63563
 ; TYPE: DNA
 ; ORGANISM: Moraxella catarrhalis
 US-10-672-787-33

Alignment Scores:
 Pred. No.: 1.42e-08 Length: 63563
 Score: 211.50 Matches: 68
 Percent Similarity: 45.07% Conservative: 60
 Best Local Similarity: 23.94% Mismatches: 111
 Query Match: 13.91% Indels: 45
 DB: 17 Gaps: 11

US-10-039-183A-2 (1-299) x US-10-672-787-33 (1-63563)
 QY 50 G1yAtgProlIeThrIysSerAspPheAspMetIleIysGlnArg-----AsnProAsn 67
 Db 11755 GGCAGACT-----TATCTGATGCTCAATTATTCAGCAAGCATTAAGTAAAGAAATT 11808
 QY 68 PheAspPheAspIysLeuIysGluIysGlnIuAlaLeuIleAspGlnAlaIleArg 87

```

Db      11809  TTTGACGACGCAATCTTATCGCCTCAAGCCACCAACTTTTACCAATAATCTAGTAAT 11868
               |||||  |||  :|||  :|||  :|||  :|||
Qy      88   ThrAlaIeuValGluSnglnAlaYsrHngJluYsLeuAspSerThrProGlu----- 105
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      11869  GGGGTACGACATTAAGACCCCAAAAGCAGTAGCTGCCGAAGTGCAGATCATCTCTCTAATAGC 11928
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      106  -----PheYsAla-----MetMetGluAlaValYsYsGln 116
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      11929  CAAAGAAAGCATTTGTTTTTATGACAGTCAAGCCTTAGTTATGTTATGCAATTAATCTGGCTCA 11988
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      117  AlaIeuValGlu-----PheTPAlaYsLeuGlnIaGluIaValYs 131
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      11899  GAACTGATTGAGTGTCTTGGCCGAGAGGGCGTGTGAGAAATGMAAGAACTACATCGCT 12048
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      132  LysVal-----GlnIleProGluYs--GluMetGlnAspPhe 143
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12049  AAGGTGTTAGAGCAGCAAGTCATCGCCAAATCTCCAGACAAAGCAAGCTGTGAGATTAC 12108
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      144  TyrAsnAlaAsnYsAspGlnIuPheValYsGlnGluAlaIaIaArgHisIleIeu 163
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12109  TATCATCTTAATCAAGTTGAAATTTGTAGATGCCAATATGACGCTTGCCATATTTTG 12168
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      164  -----ValYsThrGlnAspIuAlaYsArg 172
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12169  CTTCGCGTTCACCAACAAGAGGTGTAAGACCGCATTTGAAATTAAGAAAGCAAGCTGCCCA 12228
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      173  IleIleSerGlnIleAspYsGlnProYsAlaYsLeuGluAlaYsPheIleGluIeu 192
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12229  TTGATTCATGGGCTTAATCA-----AACCAACAATCGAATATGATTTATTTGAGTTT 12282
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      193  AlaAsnArgAspThrIleAspProAsnSerYsAsnAlaGlnAsnGlyGlyAspLeuGly 212
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12283  GCAAGTCGATACCTCGCTGCCCTTCAAAA-----GATGACATGGCCGTCAACTGGGT 12333
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      213  LysPheGlnYsAsnGlnMetAlaProAspPheSerYsAlaAlaPheAlaIeuThrPro 232
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12334  GATATTACAAAAGGACGATACCGTCTGTAATTTGAGAGTGCCGATTTGCTTG---CCA 12390
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      233  GlyAspTyrThrIleYsrHngIleProYsHngIuPheGlyTyrHisIleIleYsrIle 252
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12391  GTTGCTATCAGATTATATCGAATTGAACACGATATGGATTCAGATGTTGAAGTTCTA 12450
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      253  SerYsAspSerProValThrTyrTyrGluGlnAlaLysProThrIleYsGlyMet 272
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12451  CAAAGACGAGAGGTAGGACGAGCTGACCTTTGAAAGGCGCATTCATCTATCATGAAATCAT 12510
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      273  LeuGlnIuYsLeuPheGlnGluIaGlyMetAsnGlnArgIleGluGluIeuArgYsHis 292
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12511  CTTAAGCAGACGATTCATTATCATCTAGTTGTGTGACTATTATTTGAATTAGCCAAAG 12570
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Qy      293  AlaYsIleVal 296
               :|||:  :|||:  :|||:  :|||:  :|||:  :|||
Db      12571  GCTGATTTATT 12582

RESULT 11
US-08-781-986A-168
: Sequence 168, Application US/08781986A
: Publication No. US2003005436A1
: GENERAL INFORMATION:
: APPLICANT: Charles Kunsch
: TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
: NUMBER OF SEQUENCES: 5255
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33

```

[illegible]

```

Oy 176 GlnlleAspArgInProLysAlaLysGluAlaLysPheIleGluLeuAlaAsnArg 195
Db 4876 GAAATTCATAAAAGAA---GTTTAAAAAGATCCAAAGTAAATTTGGTGAATTCGCTAAAAA 4932
Oy 196 AspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyGlyAspLeuGlyLysPheGln 215
Db 4933 GAATCAATGATGATCGTGTTCA-----GCTAAAAAAGATGGGAAATTAGGTTATGTTCTT 4986
Oy 216 LysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeuThrProGlyAspTyr 235
Db 4987 AAAGCAACAAACGTAAAGATTTTGAAGAAAGCATTTTAAAGCTTAAAGATGGTGAAGTA 5046
Oy 236 ThrLysThrProValLysThrGluPheGlyTyrHisIleIleTyrLeuIleSerLysAsp 255
Db 5047 TCAAGAGTT---GTTAAATCAAGCTTTGGATATCATATT-----ATTAAAGCTGAT 5094
Oy 256 SerProValThrTyrThrTyrGluGlnAlaLysProThrIleLysGlyMetLeuGlnGlu 275
Db 5095 AAACCAACAGACCTTAAACAGTGA-----AAACAAAGCTGAAAGAA 5136
Oy 276 LysLeuPheGlnGluArgMetAsnGlnArgIleGluGluLeuArg-LysHisAlaLysI 295
Db 5137 AAATTAAGTCGATCAGAAAGTACAAAAAATTCAAAAATTATTAAGTCGATGATCAAAAGAT 5196
Oy 295 e 295
Db 5197 C 5197

RESULT 12
US-10-329-624-168
; Sequence 168, Application US/10329624
; Publication No. US20040043037A1
GENERAL INFORMATION:
APPLICANT: Charles Kunesh
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB Storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/329,624
FILING DATE: 27-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/956,171
FILING DATE: October 20, 1997
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,966
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark U. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 168:
SEQUENCE CHARACTERISTICS:

```

```

;          LENGTH: 7963 base pairs
;          TYPE: nucleic acid
;          STRANDEDNESS: double
;          TOPOLOGY: linear
;
;          SEQUENCE DESCRIPTION: SEQ ID NO: 168:
US-10-329-624-168

Alignment Scores:
Pred. No.:      3,186-09      Length:      7963
Score:          206.50      Matches:      84
Percent Similarity: 43.30%      Conservative: 55
Best Local Similarity: 26.17%      Mismatches: 108
Query Match:      13.59%      Indels:      74
                        Gaps:      15

US-10-039-183A-2 (1-299) x US-10-329-624-168 (1-7963)

QY      5  ILeuAaenLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMetAlaLysProAla 24
      :      :      :      :      :      :      :      :      :      :
Db      4369 ATGATTAACAATTAATTCCTCCGGTAACAGCTAGCTTTATTATTAGCGCGCTGTGAC 4428

QY      25  HisAaenAlaAaenAaenAlaThrHisAaenThrLysThrThrAaspSerSerAlaGlyVal 44
      :      :      :      :      :      :      :      :      :      :
Db      4429 GCTAGTGCACAGACTCTAAAGAAAATACA-----TTAATTTCTTTAAAGCTGAGAC 4482

QY      45  LeuAlaThrValAaspGlyAargProIleThrLysSerAaspPheAaspMetIleLysGlnArg 64
      :      :      :      :      :      :      :      :      :      :
Db      4483 GTAAACAGTTGCAGAT-----ACATGAAAAAATCCGTAAGAATCAAAATT-----GCA 4530

QY      65  AaenProAaenPheAap-----PheAapLysLeuLysGlnLysGln 77
      :      :      :      :      :      :      :      :      :      :
Db      4531 AATGATCATCTTACTGAAATGTTAAATAAATTTTAGCTGATTAATATAATAATAAAGTT 4590

QY      78  LysGlnAlaLeuIleAaspGlnAlaIleArgThrAlaLeuValGlnAaenGlnAlaLysThr 97
      :      :      :      :      :      :      :      :      :      :
Db      4591 AATGATTAAGAAGATTGACGAACAATAAT-----GAAAAAATGCAG 4629

QY      98  GlnLysLeuAaspSerThrProGluPheLysAlaMetMetGlnAlaValLysLysGlnAla 117
      :      :      :      :      :      :      :      :      :      :
Db      4630 AAGCAATACGCGCGTAAGATTAATTGAA-----AAGCCCTTCAACAGCAAGGT 4680

QY      118  LeuValGluPheThrAlaLysLysGlnAlaGlnGluValLys----- 131
      :      :      :      :      :      :      :      :      :      :
Db      4681 TTAAACA-----GCCGATTAATAATTAAGAAATTTACGTAAGCTGCTTATCATTA 4731

QY      132  -----LysValGlnIleProGluLysGlnMetGlnAaspPheLysAsnAla 146
      :      :      :      :      :      :      :      :      :      :
Db      4732 GAATTACTATCAGATTAATAATTAATCTCTGATCTGAAATTAAGAA----- 4779

QY      147  AaenLysAaspGlnLeuPheValLysGlnGlnAlaHisAlaArgHisIleLeuValLysThr 166
      :      :      :      :      :      :      :      :      :      :
Db      4780 -----GACAGCAAGAAAGCTTCACACATTTTATTAATTAAGTT 4815

QY      167  -----GluAaspGlnAlaLysArgIleIleSer 175
      :      :      :      :      :      :      :      :      :      :
Db      4816 AATCTAAGAAAAAGCGACAAAGAGAGCTTATGATTAAGAAAGCAAAAAAGCTGAA 4875

QY      176  GlnLysLeuLysGlnProLysAlaLysLysGlnAlaLysPheIleGlnLeuAlaAsnArg 195
      :      :      :      :      :      :      :      :      :      :
Db      4876 GAAATTTCAAAAAGAA--GTTTCAAAAGATCCAAAGTAATTTGTGAAATCCCTAAAAAA 4932

QY      196  AaenThrIleAaspProAaenSerLysAaenAlaGlnAaenGlyLysAaspLeuGlyLysPheGln 215
      :      :      :      :      :      :      :      :      :      :
Db      4933 GAATCAATAGATACGCTTCA-----GCTAATAAAGATGCGCAATTTAGCTTATCTTCT 4986

QY      216  LysAaenGlnMetAlaProAaspPheSerLysAlaAlaPheAlaLeuThrProGlyAaspTyr 235
      :      :      :      :      :      :      :      :      :      :
Db      4987 AAAGCAACAACATGATTAAGATTTGAAAAAGCACTATTTAAGCTTAAGATGTAAGTA 5046

QY      236  ThrLysThrProValLysThrGlnPheGlyTyrHisIleIleLysLeuIleSerLysAasp 255
      :      :      :      :      :      :      :      :      :      :
Db      5047 TCAGAGGTT---GTTAAATCAAGCTTTGATATCATATTT-----ATTAAAGCTGAT 5094
```

```

QY      256  SerProValThrTyrThrTyrGlnGlnAlaLysProThrIleLysGlyMetLeuGlnGln 275
      :      :      :      :      :      :      :      :      :      :
Db      5095 AACCAACAGACCTTTAACAGTGAA-----AAACAAAGCCTGAAAGAA 5136

QY      276  LysLeuPheGlnGlnArgMetAaenGlnArgIleGlnGlnLeuArg-LysHisAlaLysGln 295
      :      :      :      :      :      :      :      :      :      :
Db      5137 AAATTAAGTCATCAGAAAGTACAAAAAATCCAAAAATTAATTAAGTATGATGATCAACAAAGT 5196

QY      295  e 295
Db      5197 C 5197

RESULT 13
US-10-470-048B-173
; Sequence 173, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; FILE REFERENCE: SOINN.035US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 173
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-470-048B-173

Alignment Scores:
Pred. No.:      2,796-10      Length:      960
Score:          205.50      Matches:      84
Percent Similarity: 43.30%      Conservative: 55
Best Local Similarity: 26.17%      Mismatches: 108
Query Match:      13.52%      Indels:      74
                        Gaps:      15

US-10-039-183A-2 (1-299) x US-10-470-048B-173 (1-960)

QY      5  ILeuAaenLeuAlaLeuValGlyAlaLeuSerThrSerPheLeuMetAlaLysProAla 24
      :      :      :      :      :      :      :      :      :      :
Db      7  ATGATTAACAATTAATTCCTCCGGTAACAGCTAGCTTTATTATTAGCGCGCTGTGAC 66

QY      25  HisAaenAlaAaenAaenAlaThrHisAaenThrLysThrThrAaspSerSerAlaGlyVal 44
      :      :      :      :      :      :      :      :      :      :
Db      67  GCTAGTGCACAGACTCTAAAGAAAATACA-----TTAATTTCTTTAAAGCTGAGAC 120

QY      45  LeuAlaThrValAaspGlyAargProIleThrLysSerAaspPheAaspMetIleLysGlnArg 64
      :      :      :      :      :      :      :      :      :      :
Db      121 GTAAACAGTTGCAGAT-----ACATGAAAAAATCCGTAAGAATCAAAATT-----GCA 168

QY      65  AaenProAaenPheAap-----PheAapLysLeuLysGlnLysGln 77
      :      :      :      :      :      :      :      :      :      :
Db      169 AATGATCATCTTACTGAAATGTTAAATAAATTTTAGCTGATTAATATAATAAATAAAGTT 228

QY      78  LysGlnAlaLeuIleAaspGlnAlaIleArgThrAlaLeuValGlnAaenGlnAlaLysThr 97
      :      :      :      :      :      :      :      :      :      :
Db      229 AATGATTAAGAAGATTGACGAACAATAAT-----GAAAAAATGCAG 267

QY      98  GlnLysLeuAaspSerThrProGluPheLysAlaMetMetGlnAlaValLysLysGlnAla 117
      :      :      :      :      :      :      :      :      :      :
Db      268 AAGCAATACGCGCGTAAGATTAATTTGAA-----AAGCCCTTCAACAGCAAGGT 318

QY      118  LeuValGluPheThrAlaLysLysGlnAlaGlnGluValLys----- 131
      :      :      :      :      :      :      :      :      :      :
Db      319 TTAAACA-----GCCGATTAATAATTAAGAAATTTAGCTGATTAATTAATTAATTA 369

QY      132  -----LysValGlnIleProGluLysGlnMetGlnAaspPheLysAsnAla 146
      :      :      :      :      :      :      :      :      :      :
Db      370 GAATTACTATCAGATTAATAATTAATCTGATTCGAAATTAAGAA----- 417
```

Qy 147 Aenlyasbpglnleuphevalylseglnlualahlsalarghlsleleuvalylsergln 166
Db 418 -----GACAGCAAGAAAGCTTCACACTTTTAATTAAGTT 453
Qy 167 -----Gluaspglnlualylarglileleser 175
Db 454 AATCTAAGAAAAGCAGCAAGAGCTTAGATGATMAAGAGCGMAACMAAGCTGAA 513
Qy 176 Gluileasbpglnleuphevalylseglnlualahlsalarghlsleleuvalylsergln 195
Db 514 GAAATTCAGAAAAGAA---GTTTCAGAAAGATCCAGTAATTTGGTGAATTCGCTAAAAA 570
Qy 196 Aepthrlleaspproasbserlysaenlaglanaenlgllyaspleuglylysphegln 215
Db 571 GAATCATGATGATCTGCTCA-----GCTAAGAAAAGATGGGAATTAAGTTATGTTCT 624
Qy 216 Lyasbnglmetcalaproasbserlysaenlaglanaenlgllyaspleuglylysphegln 235
Db 625 AAGGACAAACTGATMAAGATTTTGGAAAAGCACTTTAAGCTTMAAGATGCTGAAGTA 684
Qy 236 Thrlysthrprovalthrtglnlualylspheglnlythrllelelyleuileserlyasb 255
Db 665 TCAGAGGTT---GTTAATCAAGCTTTGATATCATAT---ATTAAAGCTGAT 732
Qy 256 Serprovalthrtglnlualylspheglnlythrllelelyleuileserlyasb 275
Db 733 AACCACACAGACTTTAAGCTGA-----AAACAAAGCCTGAAGAA 774
Qy 276 Lyaleupheglnlualargmetasnglnarglileglnleuarglyshlsalaly 295
Db 775 AATTAATGTCGATCAGAAAGTACAAAAAATTCAAATTAATGATGATCATACAAAGAT 834
Qy 295 e 295
Db 835 C 835
RESULT 14
US-10-398-221-1250/c
; Sequence 1250, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1250
; LENGTH: 1163
; TYPE: DNA
; ORGANISM: Listeria monocytogenes-4B
US-10-398-221-1250
Alignment Scores:
Pred. No.: 1,08e-09 Length: 1163
Score: 200.50 Matches: 54
Percent Similarity: 51.53% Conservative: 30
Best Local Similarity: 33.13% Mismatches: 60
Query Match: 13.19% Indels: 19
DB: 17 Gaps: 6
US-10-039-183a-2 (1-299) x US-10-398-221-1250 (1-1163)
Qy 137 Gluylglnmetcalaproasbserlysaenlaglanaenlgllyaspleuglylysphegln 156
Db 1024 GACAAACACTTAAATAATATATGAAACATGCGCAACGATATTAAGTCTAAGC----- 971

Qy 157 Alahtsalarghlsleleuvalylserglnlualahlsalarghlsleleuvalylsergln 176
Db 970 -----CATATCTTGTAGCTGATGAAAACAAAGCCAAAGAAAGTT----- 932
Qy 177 Ileasbpglnleuphevalylseglnlualahlsalarghlsleleuvalylsergln 196
Db 931 -----GAAACAAACTGMAAGACGCGGCAAAATTTGCTGATTTAGCAAAAGAAATAC 881
Qy 197 Thrleaspproasbserlysaenlaglanaenlgllyaspleuglylysphegln 216
Db 880 TCTACAGACTGCTTCAATTAAGAC-----AATGTGCAATTAATGCTCATTTGGTCTCT 827
Qy 217 Aenglmetcalaproasbserlysaenlaglanaenlgllyaspleuglylysphegln 236
Db 826 GGTAAAGATGATCTGCTGATTTGAAAAGAGCTTAAAGCTTTAAACAAAGCGGACATC 767
Qy 237 Lysthrprovalthrtglnlualylspheglnlythrllelelyleuileserlyasb 256
Db 766 AGCGCTCCAGTMAAAACAAATACGATACCATCATCAATG-----GACAAA 716
Qy 257 Provalthrtglnlualylspheglnlythrllelelyleuileserlyasb 274
Db 715 CCGCAACAAACAACTTTCGAAAAGATMAAAAGCTGMAAAAGCTTCTTAAGCTTGA 656
Qy 275 Gluylleupheglnlualargmetasnglnarglileglnleuarglyshlsalaly 294
Db 655 TCTCAACTACTACATGACGACATGCAAAAGCCTTMAAAAGATACAAATGCTTAC 596
Qy 295 llevalile 297
Db 595 GTAAAGATG 587
RESULT 15
US-10-398-221-8
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 495269
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-8
Alignment Scores:
Pred. No.: 1.58e-05 Length: 495269
Score: 191.50 Matches: 79
Percent Similarity: 42.48% Conservative: 51
Best Local Similarity: 25.82% Mismatches: 127
Query Match: 12.60% Indels: 49
DB: 17 Gaps: 12
US-10-039-183a-2 (1-299) x US-10-398-221-8 (1-495269)
Qy 1 Metlyblysaenlaleuvalylseglnlualahlsalarghlsleleuvalylsergln 20
Db 273998 TTAATAAAGATGATGATTTCCCTA-----GTACGACGACACTACTTACTAC 274045

```
OY      21  A1aLysProAlaHisAsnAlaAsnAsnAlaThrHisAsnThrLysThrThrAspSer 40
      |||
      :|:|:|
Db 274046 GCTGGTCGCGAAGTAGTGCATAGTAATAAACAAGATGCAGAAATGTGCACACAGACGAG 274105
OY      41  SerAlaGlyValLeuAlaThrValAspGlyArgProIleThrLys----- 55
      :|:|
      :|:|
Db 274106 CTATATGAAAGCGATGAAAAAGCGATTACGGTATATGAGTCGTGCACCAACTTACTTCAAA 274165
OY      56  -----SerAspPheSerMetIleLysGlnArgAsnProAsnPheAspPheAspLys 72
      |||
      :|:|:|
Db 274166 AAAATCTTAGAAGATTAATATACCTGTACTGAAAAAGAAATGACAAATATATAAAA 274225
OY      73  LeuLysGlnLysGlnLysGlnAlaLeuIleAspGlnAlaIleArgThrAlaLeuValGlu 92
      :|:|:|
      :|:|:|
Db 274226 TACGAAGAACAAGTACGGTACATCATTT-----GAA 274255
OY      93  AsnGlnAlaLysThrGlnLysLeuAspSerThrProGluPheLysAlaMetMetGluAla 112
      :|:|
      :|:|
Db 274256 TCACCTTATCGTCCATTAATTTTAACAAAACCT--TCTTCMAA-----GAAAAAT 274303
OY      113  ValLysLysGlnAlaLeuValGluPheThrAlaLysLysGlnAlaGluGluValLysLys 132
      :|:|:|
      :|:|:|
Db 274304 TTACAGCTATACCTTCTAGTTCA-----AAGCAACAGAACCA--AAT 274345
OY      133  ValGlnIleProGluLysGlnLysMetGlnAspPheTyrAsnAlaAsnLysAspGlnLeuPhe 152
      :|:|
      :|:|:|
Db 274346 ATGAATGTAGAGTAAGTAATACTAATAAACTTACAAAACCTTGGAGCCAAACATCACT 274405
OY      153  ValLysGlnGluAlaHisAlaArgHisIleLeuValLysThrGluAspGluAlaLysArg 172
      |||
      :|:|:|
Db 274406 GTG-----CGTCATATTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 274447
OY      173  IleIleSerGluIleAspLysGlnProLysAlaLysLysGluAlaLysPheIleGluLeu 192
      |||
      :|:|
Db 274448 ATT-----CAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 274489
OY      193  AlaAsnArgAspThrIleAspProAsnSerLysAsnAlaGlnAsnGlyGlyAspLeuGly 212
      |||
      :|:|
Db 274490 GCTAAAGAGTACTCAACAGAT-----ACTGCACACTAGCACAAATGGCGACTATTAGAT 274543
OY      213  LysPheGlnLysAsnGlnMetAlaProAspPheSerLysAlaAlaPheAlaLeuThrPro 232
      :|:|
      :|:|
Db 274544 CCATTGTGCTCGTGGAAATGATGAAACTTTTGAAAAAGCTGCTTATGCACTTAATAAAT 274603
OY      233  GlyAspTyrThrLysThrProValLysThrGluPheGlyTyrHisIleIleTyrLeuIle 252
      |||
      :|:|:|
Db 274604 AAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 274663
OY      253  SerLysAspSerProValThrTyrThrTyrGluGlnAlaLysProThrIleLys--Gly 271
      :|:|
      :|:|
Db 274664 AAGAAAAACGAAAAAGGACCTTATGCAAAAGAAAAAGCA-----ACGTAAAAAGCAGCC 274717
OY      272  MetLeuGlnGluLysLeuPheGlnGluArgMetAsnGlnArgIleGluGluLeuArgLys 291
      :|:|:|
      :|:|:|
Db 274718 TATATTAATCTCAATTAATCTTCTGAATAATATGACAGCTGCCCTTAATAAAAGAAATTAATA 274777
OY      292  HisAlaLysIleValIle 297
      |||
      :|:|
Db 274778 GCACTTAATATGATATTT 274795
```

Search completed: April 24, 2005, 12:07:29
Job time : 1665.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2005, 05:28:36 ; Search time 3354.11 Seconds
(without alignments)
3393.213 Million cell updates/sec

Title: US-10-039-183A-2

Perfect score: 1520

Sequence: 1 MKKQILNLTALVGLSTSLFM.....ERMQRPELRKRAKIVINK 299

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US10039183/runat_22042005_122425_20124/app_query.fasta_1.1038
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10039183@cgn2.1_1.5533@runat_22042005_122425_20124 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: gb_eest1:*
2: gb_eest2:*
3: gb_hic:*
4: gb_eest3:*
5: gb_eest4:*
6: gb_eest5:*
7: gb_eest6:*
8: gb_gsest1:*
9: gb_gsest2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	229	15.1	701	BH369380	AG-ND-155
2	217.5	14.3	705	AQ242234	4P12-64F
3	215.5	14.2	2256	CL982381	OsIRSC047
4	203.5	13.4	669	AG366881	PingSC047
5	195.5	12.9	2850	CL982316	OsIRSC047
6	178.5	11.7	1203	BZ549176	pacel-60
7	177.5	11.7	543	CF842872	peH8022xK
8	169.5	11.2	666	BE776559	MY-17-B-0
9	161	10.6	629	BE776294	MY-14-A-1

10	160	10.5	672	2	BE776554	MY-17-B-0
C 11	159.5	10.5	658	9	AG613321	Escherich
12	159.5	10.5	774	9	CL680822	PR1012b H
C 13	158.5	10.4	972	8	BZ568565	pacel-154
14	152.5	10.0	1540	8	BZ574276	mh2_3590
C 15	151.5	10.0	502	7	CF843490	pH8025xG
16	150	9.9	429	7	CF587341	rv56b05.Y
C 17	148.5	9.8	543	1	AV907878	AV907878
18	148.5	9.8	422	7	BE034276	MH02E09 M
C 19	148	9.7	482	7	CF101411	rv18g04.Y
C 20	146.5	9.6	754	6	CB821785	EST_2636
21	145.5	9.6	563	5	BP013190	BP013190
C 22	145	9.5	564	1	AV854521	AV854521
23	145	9.5	604	2	AV966480	AV966480
24	144.5	9.5	731	1	AU003742	AU003742
C 25	144	9.5	652	5	BM268878	BM268878
C 26	144	9.5	1036	8	BZ577220	pacel-60
27	143.5	9.4	355	7	CF856325	pHML007xM
28	143.5	9.4	660	1	AU000274	AU000274
29	143.5	9.4	722	1	AU003857	AU003857
30	143.5	9.4	729	1	AU000273	AU000273
31	143.5	9.4	738	1	AV400559	AV400559
32	143.5	9.4	740	1	AU003792	AU003792
33	143.5	9.4	748	1	AU003598	AU003598
34	143.5	9.4	750	5	BP115577	BP115577
35	143	9.4	554	2	AV976003	AV976003
36	143	9.4	661	7	CK752673	pam01-2ms
37	142.5	9.4	608	6	CB830929	BN40_047F
38	142.5	9.4	699	1	A1822163	A1822163
39	142	9.3	360	1	AV198267	AV198267
40	142	9.3	420	4	BJ754894	BJ754894
41	142	9.3	507	4	BJ761135	BJ761135
42	142	9.3	515	7	CF828017	Mg_AFB_06
43	142	9.3	573	2	AV996173	AV996173
44	142	9.3	573	7	CN088091	EC2BBA29C
45	142	9.3	656	7	CN657859	rt55c06.Y

ALIGNMENTS

RESULT 1
BH369380/c
LOCUS
DEFINITION AG-ND-155G10.TF ND-TAM Anopheles gambiae genomic clone
ACCESSION AG-ND-155G10, genomic survey sequence.
VERSION BH369380.1 GI:17315481
KEYWORDS
SOURCE
ORGANISM
Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 701)
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J.,
Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B.,
Gardner, M.J. and Collins, F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito *Anopheles gambiae*
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
JOURNAL
MEDLINE
PUBMED
COMMENT
Other_GSSs: AG-ND-155G10.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research

(TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For
Class: BAC ends.

Class: BAC

FEATURES

```
Location/Qualifiers
1..701
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PBR"
/db_xref="taxon:7165"
/clone="AG-ND-15G10"
/clone_1b="ND-TM"
note="vector: pCBBC1; Site 1: HindIII"
```

ORIGIN

Alignment Scores:	
Pred. No.:	1.67e-15
Score:	229.00
Percent Similarity:	48.36%
Best Local Similarity:	34.47%
Query Match:	15.07%
DB:	8
	length: 701
	Matches: 73
	Conservative: 30
	Mismatches: 76
	Indels: 9
	Gaps: 3

US-10-039-183A-2 (1-299) x BH369380 (1-701)

Qy	8	ArgThrAlaIeuValGluGlnGlnuValLysThrGluLysIleuAspSerThrProGluPhe	106
Db	682	AGGCGGGCGCTGGCCCAACAGCCGAGCC-----TAC	653
Qy	107	LysAlaMetGluValIaValLysGlnAlaIeuValGluPheTrpAlaLysGln	126
Db	652	AAAGAGACAGATGGAACTGGCGCGCAGACCATCTGCATC-----CGCTCCGTG	605
Qy	127	AlaGluGluVal---LysLysValGlnIleProGluLysGluMetGlnAspPheTyrAsn	145
Db	604	TTGAGAGACTTCCAGAAAGAAACCCCGTAGCCGACTGAAGCCAA-----	557
Qy	146	AlaAsnLysAspGlnIleuPheValLys-----GlnGluValIaIaLysGlnIle	162
Db	556	GCCGATACGACAAAGCCCGCTCGTGCACACGCGCAGAGATACAAAGCCAGCAATC	497
Qy	163	LeuValLysThrGluAspGluValaLysArgIleIleSerGlnIleAspLysGlnProLys	182
Db	496	CTGGTCGAGTCGGAAAGACCGCGCAAGGCATCATTCGCGAGATC-----AAG	449
Qy	183	AlaLysLysGlnLysPheIleGluIleuValaAsnArgAspThrIleAspProAsnSer	202
Db	448	GCCGGCAAG-----AAGTTGAAAGACATGCGCAAGAAAGAGTCCAAAGATCCCGGATCC	395
Qy	203	LysAsnAlaGlnAsnGlyLysAspLeuGlyLysPheGlnLysAsnGlnMetAlaProAsp	222
Db	394	-----GGCGCCCGTGGCGGTGACCTCGACTGGGCAACCCCGCACTACGTCCCGAG	341
Qy	223	PheSerLysAlaIaIaPheAlaLeuThrProGlyAspTyrThrLysThrProValLysThr	242
Db	340	TTCTCCGAACCCCTGTATCAAGCTCGAAAGGGGTGGCATAGACCCAGAACCCGCTCAAGACC	281
Qy	243	GluPheGlyTyrHisIleIleTyrLeuIleSerLysAspSerProValThrTyrThrTyr	262
Db	280	CAGTTGGGCTACACATCATCTCGCTCGACGATTCACGCCAGCGCGAGCTCCCAAGTTC	221
Qy	263	GluGlnAlaLysProThrIleLysGlyMetLeuGlnGluLysLeuPheGlnGluArgMet	282
Db	220	GAGGAAGTCAAGCCCAAGATCGGACGACGACTGCAA-----CAGCAAAAGCTG	173
Qy	283	AsnGlnArgIleGluGluLeuArgLysHisAlaLysIle	295
Db	172	GCCCAATTCAGGAATCGCTGGCGCAAAAGGCCAAGATC	134

FEATURES

Source

High quality sequence stop: 1.

1. .705

1. .705

```
/organism="Ochrobactrum anthropi"  
/mol_type="genomic DNA"
```

```
/lab_host="DH10B E. coli"  
clone_lib="Ochrobactrum anthroni BAC library"
```

/note=Vector: pBeloBAC11, Site:1: HindIII; Site 2:
 HindIII; Ochrobactrum anthropi is an important microbe
 having potential for the bioremediation of environments
 contaminated by aliphatic compounds. We have constructed
 a BAC library for *O. anthropi* that provides a 90x genome
 coverage based on an estimated genome size of 3.83 Mb.
 The library contains 3072 clones with an average insert
 size of 112 kbp. High-density colony filters of the
 library were made and a physical map of the genome
 constructed using a hybridization without replacement
 strategy. In addition, 624 randomly chosen BAC clones
 were HindIII fingerprinted and analyzed using Fingerprint
 coding (FPC; Sanger Centre, UK). The FPC results closely
 supported and verified the hybridization contig data.
 After determining a reduced tiling path of 69 clones,
 138 BAC ends were sequenced for a genome wide survey of
 gene distribution and gene structure. "

ORIGIN

Alignment Scores:	
Pred. No.:	3 43e-14
Score:	217.50
Percent Similarity:	61.47%
Best Local Similarity:	42.20%
Query Match:	14.31%
DB:	8
Length:	705
Matches:	46
Conservative:	21
Mismatches:	35
Indels:	7
Gaps:	2

US-10-039-183A-2 (1-299) X AQ242234 (1-705)

```

QY      187  ALALysPheH1EG1LUEuA1AsnAtrgspHr1LeAsPProAAsnSerLySAAsn1AGln 206
          :::::
Db       8   GCCAAATTCGAGGACTCGCCAGAGAGACTTCAACCGAC-----GGCACCGCGAGCC 58
          :::::
QY      207  AAsnG1yG1AspLeuG1yLysPheG1nLySAAsnG1nMeA1AProAspPheSerLySA1a 226
          :::::

```

Db 59 AATGGCGCATCTCGCTATTTCGCCGAGCCAGCATGTCCGAGTTCCGAAAAAGCC 118
 Qy 227 AAlphealeuThProGlyAapTyThrlyThrProVallyThrgluPheGlyTy 246
 Db 119 GCCTTGGCTCAAGCCGGGTGAATACACCAAGAGCCGCTCAAGCCAGTTCCGCTTC 178
 Qy 247 HsilelleTyLeulleseerlysaPseerProvalThrTyThrTygluInlaTy 266
 Db 179 CATGTATCCAGCTCGAAGATGTGCGACCAAGCCAGCCGCGATTCATCATGATCAAG 238
 Qy 267 ProthrIlelysglyMetleuInglulysleuPheIngluTyxMetAsnInly 286
 Db 239 GACCATCGCTTCATC-----ATCATGCGGAGCGTTATGTGAACCCGTC 286
 Qy 287 GluGluLeuArglyHsilelle 295
 Db 287 AAGAAGCTGCTGACGATGATGAATC 313
 RESULT 3
 LOCUS CL982381 2256 bp DNA linear GSS 21-SEP-2004
 DEFINITION OsJFSC047527 Oryza sativa Expressed Library Oryza sativa (indica
 ACCESSION CL982381
 VERSION CL982381.1 GI:52419244
 KEYWORDS GSS.
 ORGANISM Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 2256)
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wang, G. K. S., Deng, X. W. and Wang, J.
 An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 Unpublished (2004)
 JOURNAL
 COMMENT Contact: Chen Chen
 Department of Bioinformatic
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80486576
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 FEATURES.
 source 1..2256
 location/Qualifiers
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed Library"
 /note="Oryza sativa exon trapped genomic sequences"
 ORIGIN
 Alignment Scores: 2.78e-13 Length: 2256
 Score: 215.50 Matches: 64
 Percent Similarity: 52.11% Conservative: 47
 Best Local Similarity: 30.05% Mismatches: 92
 Query Match: 14.18% Indels: 10
 Db: 9 Gaps: 7
 US-10-039-183a-2 (1-299) x CL982381 (1-2256)
 Qy 87 ArgThraAlaLeuValGluAsnGluAlaValysThrGluLeuAspSerThrProGluPhe 106
 Db 795 GCGCTGCTCATGGCTCA---CGCGTTGAAGCCGACGATTCGAGAAGACCCGAAAGTC 851
 Qy 107 LysAlaMetMetGluAlaValAlaValysGlnAlaLeuVal-Glu---PheTrpAlaLys 125
 Db 852 GCCGCTGATTCGGCGTGGCCGCGAGCAAGGTGCTGTCCAGCGCTATCTGTCAAGGT 911

Qy 125 sglAlaGluValValysValGlnIleProGluTyGluMetGlnAspPheTy 145
 Db 912 GGACAGAGACACATGCTTTGGAGCTCGGCTCCGAAAGGTCTGGCTCGCAACATCTATA 971
 Qy 145 nAlaAsnlysaPseerInleuPheVallyleGlnIuAlaHsilaArgHsilelleVally 165
 Db 972 GGCAGAGCTGGAAGCTTCAGAGGTCTGAGAGGTGAGGTGGCGCATCTGTTC 1031
 Qy 165 s---ThrgluPheGlnAlaValArgIlelleseerIulleAspTyGlnProlyAla 184
 Db 1032 AGGCAGGGGGGGGCTCCCGGCGCCAGATCGAAGATCGAACGAACTCAGCA-- 1089
 Qy 184 slyleGluAlaTyPheIleGluAlaAsnArgAspThrIleAspProAsnSerTy 204
 Db 1090 ---GCGGACATTTGCGACAGCTGCGCAAGAGCGCTCTGCGCAAGCGCAT- 1140
 Qy 204 nAlaGlnAsnGlyIlePheLeuGlyLyPheGlnlysaPseerProvalThrTyThrTygl 224
 Db 1141 -GCTCGAAGAGCGCGCATCTGGCTTTGAAACAGGAAGATGTGCCAGATTGCA 1199
 Qy 224 rlyAlaAlaPheAlaLeuThr---ProGlyAapTyThrlyThrProVallyThrGl 243
 Db 1200 CAAAGCTCGTTGCGCATGACCAAGCTGCGAGATTGAGCGGCATC--GTGAGAGCA 1256
 Qy 243 uPheGlyTyThrHsilelleTyLeulleseerlysaPseerProvalThrTyThrTygl 263
 Db 1257 GTTGGCTTACCATCTGCACTGCAATCGAAGCGCCCATCCCGAGAAACGCGCATTTCA 1316
 Qy 263 uGlnAlaTyProthrIlelysglyMetleuInglulysleuPheIngluTyxMetAsn 283
 Db 1317 TGAGGTGGTGTGTGATGATGACAGAGATCGCGCGCAAGCTGACAGAGAGCCAGGT 1376
 Qy 283 nGlnArgIleGluGluLeuAspGlyHsilelle 295
 Db 1377 TGCCGATGCCAGCGCTGACAGAGAGAGAGAGGATC 1413
 RESULT 4
 LOCUS AG266881 669 bp DNA linear GSS 06-AUG-2004
 DEFINITION Finegoldia magna DNA, clone: BF29r, genomic survey sequence.
 ACCESSION AG266881
 VERSION AG266881.1 GI:38175346
 KEYWORDS GSS.
 ORGANISM Finegoldia magna
 Finegoldia magna
 Bacteria; Firmicutes; Clostridia; Clostridiales;
 Peptostreptococaceae; Finegoldia.
 1 (bases 1 to 669)
 Goto, T., Todo, K., Miyamoto, K. and Akimoro, S.
 Bacterial artificial chromosome library of Finegoldia magna ATCC
 29328 for genetic mapping and comparative genomics
 Microbiol. Immunol. 47 (12), 1005-1016 (2003)
 JOURNAL
 PUBMED 14695451
 REFERENCE 2 (bases 1 to 669)
 AUTHORS Goto, T., Todo, K., Miyamoto, K. and Akimoro, S.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUN-2003) TakaraGnu Goto, Wakayama Medical
 University, Department of Microbiology; Kimidaira 811-1, Wakayama,
 Wakayama 641-0012, Japan (E-mail: t-goto@wakayama-med.ac.jp,
 Tel: 81-73-441-0640, Fax: 81-73-448-1026)
 FEATURES
 source 1..669
 location/Qualifiers
 /organism="Finegoldia magna"
 /mol_type="genomic DNA"
 /strain="ATCC29328"
 /db_xref="taxon:1260"
 /clone_lib="BF29r"
 /clone_lib="F. magna ATCC29328 wakayama BAC"
 /note="putative peptidyl-prolyl cis-trans isomerase"
 ORIGIN
 Alignment Scores:


```

Db      688 CTGCAGCTGGAGAGCCCGCAATCCCTCCAAAGACGCCGACGTTGCAGAGAGTGGC----- 681
Qy      269 lletyglwmetleuglglulvyleupheingluargmetasnglnarglleglu 288
Db      682 -----ATGCACTGATCCAGAG 699
Qy      289 leuatyglvshsalalys 294
Db      700 ATCCGCGCTGAATGCCAG 717

RESULT 6
BZ549176/c
LOCUS      BZ549176                1203 bp    DNA        linear    GSS 17-DEC-2007
DEFINITION pacel1-60_1717.61 pacel1-60 Pseudomonas aeruginosa genomic clone
ACCESSION  BZ549176
VERSION    BZ549176
KEYWORDS   BZ549176.1 GI:27152757
SOURCE     GSS.
ORGANISM   Pseudomonas aeruginosa
            Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1 (bases 1 to 1203)
            Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
            Burns, J.L., Kaul, R. and Olsen, M.V.
            Whole-genome-sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
            J. Bacteriol. (2002) In press
            Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: ckraymond@u.washington.edu
            Class: shotgun.
            Location/Qualifiers
                source          1..1203
                               /organism="Pseudomonas aeruginosa"
                               /mol_type="genomic DNA"
                               /strain="1-60"
                               /db_xref="taxon:287"
                               /clone="pacel1-60_1717"
                               /clone1ib="pacel1-60"
                               /note="Clinical isolate 1-60 whole genomic shotgun
                               library."

FEATURES
    source          1..1203
    Location/Qualifiers
        Alignment Scores:
            Pred. No.:      1.95e-09      Length:      1203
            Score:          178.50         Matches:      65
            Percent Similarity: 46.41%      Conservative: 45
            Best Local Similarity: 27.43%    Mismatches:  92
            Query Match:    11.74%         Indels:      35
            Gaps:           8              Gaps:        9

US-10-039-183A-2 (1-299) x BZ549176 (1-1203)
Qy      63 Glnatgaanprrcshsrpheasphespheslysglulvsglulys-Glnalaleu1 82
Db      916 CAAAGAAAGCC-----GACTTTGAACCGACCGAAGCAAGCTCGAAGAGCCCTA-- 865
Qy      82 eaepglnalaleatgtrhAlaleuValgluabnglu1alysrthnglulvyleuaps 102
Db      864 -----CTTACGAGAGAGCCCGACCAAGAGAGCCGAAATTCATTG 827
Qy      102 rThrProgluphelysalametmetclu--AlaVallyblysglnalaleuValgluph 121
Db      826 ACTTCGCGAGACCAAGTTGCTTCGTCGAATTAACGTTGAGCGTGAAGAAATTCCTTCCTT 767
Qy      121 eTrrpalalyblysgln-----AlsglulgluVallyblyblyValglnlleProglulysgl 139

```

Df		766	TTTGACCAAGGCTGAAGGTAAAGCACAAGAAAGACTTCGGAGCGGCTTACAGAGAAGA	707
Oy		139	UmetGlnAerPhyTyArvAlaAvnLyVaBrGlIneuPhueVallywGIngluaIAnIwaI	159
Df		706	AATGCCAACCTTTTCCGAGCA-----GGCGCATG	677
Oy		159	aArx-HisIleLeuVallyThrGlnAer-----GluAlaLyvaRgIleI	174
Df		676	CCGCACACATCTTAATGATCGAGTGAACAGAACAGTCGGCTACAGCACAGGCCCAAGGCAAGA	617
Oy		174	IeserGluIleAerLyGlnProLyvaIalyVlywGluAlalyAerPheIlegIueuAaa	194
Df		616	TCCAGAGATC-----AAGGCTCGCGCTCGGCGCAAGGCGAGAGATTTCGCGCGCTGGCA	563
Oy		194	vnaIgaBrThrlIAeRProAvnSerLyvaAvnIagIAnenGIyIAvRLeuGIyLyeR	214
Df		562	AGAGATTCTCCAGAGATAATCGGCTCG-----GCCGCACCGCGCGGCTGACCTTGCTGACG	509
Oy		214	heGInLyvaAvnGImetAlaProAvRPheseRlyvaIaAlaPhaAlaIeuThrProGIJa	234
Df		508	CCGATCGCGGCGGTACAGACCCCGCGTTCGAGAGCGGCTGTATGCGCTGAAGCAAGGTG	449
Oy		234	sPTyThlLyThrProValLyThrGlnPhueGInyTyrHIsIleIetyReuIleserI	254
Df		448	AG---GATACGCGCCCGGTGAANAkSTCGSTACGGCTACACTATCAAGCTGCTGGCG	392
Oy		254	yVaerSerProValThrTyThrTyTyrGIngluaIalyRProThrlIeLywGlyMetLeug	274
Df		391	TGCAGCGCGCGGAGTAGTACCAGCGCTGGAAGACCTCAAGCCGAGACCTCGAGAGCAACTGA	332
Oy		274	InGIuLyVleuPhueGInIuAygrMeKbnGInAgIIegIuGInIeu	289
Df		331	AGAAACAGATGTGAGACAGCGCTTCTGTCAGGCTTACCAAGAGACTG	285

RESULT 7
CF842872/c

LOCUS CF842872 543 bp mRNA linear EST 30-OCT-2003

DEFINITION penB022XK01 USDA-IPARS: Expression of Phytophthora sojae genes during infection and propagation_ahb Phytophthora sojae cDNA clone SHB022K01 5, mRNA sequence.

ACCESSION CF842872

VERSION CF842872.1 GI:38058526

KEYWORDS EST.

SOURCE Phytophthora sojae

ORGANISM Phytophthora sojae

REFERENCE 1 (bases 1 to 543)

AUTHORS Tyler,B.

TITLE Tyler,B. Not Published

JOURNAL Unpublished (2003)

COMMENT Contact: Tyler B

 Tyler lab

VARI

1860 Pract'Dr., Blacksburg, VA 24061, USA

Tel.: 540-231-7318

Email: bmttyler@vt.edu

PCR primers

FORWARD: BK reverse primer

BACKWARD: BK reverse primer

Plate: 022 row: K column: 01

Seq primer: BK reverse primer

High quality sequence stop: 543.

Location/Qualifiers

1..543

/organism="Phytophthora sojae"

/mol_type="mRNA"

/db_xref="taxon:67593"

/clone="SHB022K01"

/tissue type="mycelium"

/cell_line="pe497"

/dev_stage="48 hr. post infection stage"

/lab_host="Soybean plant"

ORIGIN

/clone_lib="USDA-IPAFS:Expression of Phytophthora sojae
genes during infection and propagation SH8"
/note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

Alignment Scores:

Pred. No.:	8.69e-10	Length:	543
Score:	177.50	Matches:	42
Percent Similarity:	55.45%	Conservative:	14
Best Local Similarity:	41.58%	Mismatches:	38
Query Match:	7	Indels:	7
DB:	7	Gaps:	3

US-10-039-183a-2 (1-299) x CF842872 (1-543)

QY 154 LysGInGluAlaHisIleuValIysThrGluAspGluAlaLysArgIleIleSerGlu 173
DB 449 AACCCGACGGCGCGTCCGACCATCTGTGCGCAGAGAGAGAGTCCAGCAGATC 390
QY 174 ILeSerGluIleAspIysGInProLysAlaLysIysGluAlaLysPheIleGluLeuAla 193
DB 389 TTGAAGAGAGCTG-----CAGCGGGCCGACGACCTGGAGCGACGTTGGCGGGCTGGCG 336
QY 194 AsnArgAspThrIleAspProAsnSerIysAsnAlaGlnAsnGlyGlyAspLeuGlyLys 213
DB 335 AAGGACGCGACGACAGTGTCTCGTCCAGTCCG-----AAGGAGGGGACCTGGGGTCC 285
QY 214 PheGInLysAsnGluMetAlaProAspPheSerIysAlaAlaPheAlaLeuThrProGly 233
DB 284 TTCGGTGGCGGCGCATGTGTGCGGATTCGACAGGTGGCGTTCAGAGAGCCCGTGGG 225
QY 234 AspTyrThrLysThrProValIysThrGluPheGlyTyrHisIleIleTyrIleuIleSer 253
DB 224 GACGTCCACAG-----GTCAAGACGCGTTTGATGCGCACTTGCTGATCAACGAC 171
QY 254 Lys 254
DB 170 CGC 168

RESULT 8

LOCUS BE776559 666 bp mRNA linear EST 20-SEP-2000
DEFINITION MY-17-8-09 PinfectansMY Phytophthora infestans cDNA, mRNA sequence.
ACCESSION BE776559
VERSION BE776559.1 GI:10230214
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

REFERENCE 1 (bases 1 to 666)
AUTHORS Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
TITLE Initial assessment of gene diversity for the complete pathogen
JOURNAL Phytophthora infestans based on expressed sequences
MEDLINE Fungal Genet. Biol. 28 (2), 94-106 (1999)
PUBMED 20056376
COMMENT Contact: Govers F
Laboratory of Phytopathology
Wageningen University
Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES

source 1..666
location/Qualifiers
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DHS-alpha"

ORIGIN

/clone_lib="PinfectansMY"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Total
RNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."

Alignment Scores:

Pred. No.:	9.33e-09	Length:	666
Score:	169.50	Matches:	39
Percent Similarity:	57.61%	Conservative:	14
Best Local Similarity:	42.39%	Mismatches:	32
Query Match:	11.15%	Indels:	7
DB:	2	Gaps:	3

US-10-039-183a-2 (1-299) x BE776559 (1-666)

QY 157 AlaHisAlaArgHisIleuValIysThrGluAspGluAlaLysArgIleIleSerGlu 176
DB 410 GCCCGCGGTTCCATATCTGTGTAATCTGAGACGAGGCTGATTAATTGTTAAGAG 469
QY 177 ILeAspLysGInProLysAlaLysIysGluAlaLysPheIleGluLeuAlaAsnArgAsp 196
DB 470 ATCGAC-----GCTGCCGAGACAGAGAGCAATGTCGGAACCTGGAGAAAGC 523
QY 197 ThrIleAspProAsnSerIysAsnAlaGlnAsnGlyGlyAspLeuGlyLysPheGInLys 216
DB 544 AGCACTTGTCTCGTCCGACAG-----AAGGAGAGAGACTGGGACATGTTCCGACGT 574
QY 217 AsnGlnMetAlaProAspPheSerIysAlaAlaPheAlaLeuThrProGlyAspTyrThr 236
DB 575 GCGGAGATGGATCCGACATTTGACAAAGTGGTGTGAAGAGAGAGTGGCGAGCTCGCC 634
QY 237 LysThrProValIysThrGluPheGlyTyrHisIle 248
DB 635 AAG-----GTGCAGACCCAGTTCGGTGGCATGTG 664

RESULT 9

LOCUS BE776294 629 bp mRNA linear EST 20-SEP-2000
DEFINITION MY-14-8-10 PinfectansMY Phytophthora infestans cDNA, mRNA sequence.
ACCESSION BE776294
VERSION BE776294.1 GI:10229949
KEYWORDS EST.
SOURCE Phytophthora infestans (potato late blight agent)
ORGANISM Phytophthora infestans
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

REFERENCE 1 (bases 1 to 629)
AUTHORS Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
TITLE Initial assessment of gene diversity for the complete pathogen
JOURNAL Phytophthora infestans based on expressed sequences
MEDLINE Fungal Genet. Biol. 28 (2), 94-106 (1999)
PUBMED 20056376
COMMENT Contact: Govers F
Laboratory of Phytopathology
Wageningen University
Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands
Tel: 31 317 483 138
Fax: 31 317 483 412
Email: Francine.Govers@medew.fyto.wau.nl.

FEATURES

source 1..629
location/Qualifiers
/organism="Phytophthora infestans"
/mol_type="mRNA"
/strain="DDR7602, A1 mating type"
/db_xref="taxon:4787"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DHS-alpha"


```

/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
/clone="454-4P"
/Note="This sequence is an updated part of the sequence
from AE000111-AE000510 series. The first three characters
of the clone name correspond to x of AE000xxx."

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1,266-07 Length: 658
Score: 159.50 Matches: 48
Percent Similarity: 51.1% Conservative: 21
Best Local Similarity: 35.56% Mismatches: 33
Query Match: 10.49% Indels: 33
DB: 9 Gaps: 6

```

US-10-039-183a-2 (1-299) x AG613321 (1-658)

```

QY 150 GlnLeuPheValLys-----GlnGluAlaHisAlaArgHisIleLeuVal 164
   |||||
DB 495 CAATTATTCACCTCAAGGCGATCATGCGCAAAACAGACGACGACATGATCTTCTGA 436
   |||||
QY 165 LysThrGluAapGluAlaLysArgIleIleSerGluIleAapLysGlnProLysAlaLys 184
   |||||
DB 435 AAAGAAGAAACTGGCTCTGATCTTCCGACGACATT-----AAG 394
   |||||
QY 185 LysGluAlaLysPheIleGluLeuAlaAsnArgAapThrIleAapProAsnSerLysAsn 204
   |||||
DB 393 AACGGGGCGATTTCCGCAAGCTGGCAAGAAACACTCCATTGGCCCATCAGCGCAA--- 337
   |||||
QY 205 AlaGlnAaGlyGlyAapLeuGlyLysPheGlnLysAsnGlnMetAlaProAapPheSer 224
   |||||
DB 336 -----CCGGGGGATTTAGTGAATTCGCCAGGGTCAGATGCTTCCGCTTCAT 283
   |||||
QY 225 LysAlaAlaPheAla-----LeuThrProGlyAapTyThrLysThrProValLys 241
   |||||
DB 282 AAAGGCTTTCTCTGTCGCGTACTGAGCCG-----ACCGCCCGCTGCAC 235
   |||||
QY 242 ThrGluPheGlyTyThrIleIleTyLeuIleSerLysAapSerProValThrTyThr 261
   |||||
DB 234 ACCGATTCGATATCATCATTAAGGTGCTGTACCGCAAC----- 193
   |||||
QY 262 TyrGluGlnAlaLysProThrIleLysGlyMetLeuGlnGluLys 276
   |||||
DB 192 -----TAATAGCAAGGCTTCTCCAGAGAGAG 166
   |||||

```

RESULT 12
CL680822 774 bp DNA linear GSS 09-JUL-2004
LOCUS PR1012b_H01_2 - PR1012b.BR (774) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
CL680822
ACCESSION CL680822.1 GI:50187782
VERSION GSS
KEYWORDS Pristionchus pacificus
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Bukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 774)
REFERENCE Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AUTHORS AppDB: an Acedb database for the nematode satellite organism
TITLE Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
JOURNAL Contact: Sommer RJ
COMMENT Max-Planck-Institute for Developmental Biology
Speaninstr. 37-39, Tuebingen D-72076, Germany
Tel: 004970701601371
Fax: 004970701601498

```

Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: 77
Class: fosmid ends.
Location/Qualifiers

```

FEATURES

```

1..774
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_id="Mixed stage fosmid library of P. pacificus
var. California"
/Note="Vector: pBplfos-5 Fosmid vector"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1,576-07 Length: 774
Score: 159.50 Matches: 48
Percent Similarity: 51.1% Conservative: 21
Best Local Similarity: 35.56% Mismatches: 33
Query Match: 10.49% Indels: 33
DB: 9 Gaps: 6

```

US-10-039-183a-2 (1-299) x CL680822 (1-774)

```

QY 150 GlnLeuPheValLys-----GlnGluAlaHisAlaArgHisIleLeuVal 164
   |||||
DB 160 CAATTATTCACCTCAAGGCGATCATGCGCAAAACAGACGACGACATGATCTTCTGA 219
   |||||
QY 165 LysThrGluAapGluAlaLysArgIleIleSerGluIleAapLysGlnProLysAlaLys 184
   |||||
DB 220 AAAGAAGAAACTGGCTCTGATCTTCCGACGACATT-----AAG 261
   |||||
QY 185 LysGluAlaLysPheIleGluLeuAlaAsnArgAapThrIleAapProAsnSerLysAsn 204
   |||||
DB 262 AACGGGGCGATTTCCGCAAGCTGGCAAGAAACACTCCATTGGCCCATCAGCGCAA--- 318
   |||||
QY 205 AlaGlnAaGlyGlyAapLeuGlyLysPheGlnLysAsnGlnMetAlaProAapPheSer 224
   |||||
DB 319 -----CCGGGGGATTTAGTGAATTCGCCAGGGTCAGATGCTTCCGCTTCAT 372
   |||||
QY 225 LysAlaAlaPheAla-----LeuThrProGlyAapTyThrLysThrProValLys 241
   |||||
DB 373 AAAGGCTTTCTCTGTCGCGTACTGAGCCG-----ACCGCCCGCTGCAC 420
   |||||
QY 242 ThrGluPheGlyTyThrIleIleTyLeuIleSerLysAapSerProValThrTyThr 261
   |||||
DB 421 ACCGATTCGATATCATCATTAAGGTGCTGTACCGCAAC----- 462
   |||||
QY 262 TyrGluGlnAlaLysProThrIleLysGlyMetLeuGlnGluLys 276
   |||||
DB 463 -----TAATAGCAAGGCTTCTCCAGAGAGAG 489
   |||||

```

RESULT 13
BZ568565 972 bp DNA linear GSS 17-DEC-2002
LOCUS pac62-164_7593.y2 pac62-164 Pseudomonas aeruginosa genomic clone
pac62-164_7593, genomic survey sequence.
BZ568565
ACCESSION BZ568565.1 GI:27201734
VERSION GSS
KEYWORDS Pseudomonas aeruginosa
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 972)
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
AUTHORS Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
JOURNAL Contact: Chris K. Raymond
COMMENT

Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES

Location/Qualifiers
1..972
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_7593"
/clone_1ib="pacs2-164"
/note="Clinical isolate 2-164 whole genomic shotgun library."

ORIGIN

Alignment Scores:
Pred. No.: 2,786-07 Length: 972
Score: 158.50 Matches: 59
Percent Similarity: 44.98% Conservative: 35
Best Local Similarity: 28.23% Mismatches: 92
Query Match: 10.43% Indels: 24
DB: 8 Gaps: 7

US-10-039-183a-2 (1-299) x BZ568565 (1-972)

QY 95 AAlaYrthGlnUlyLeuApsrThrProGluPhelYsAlaMetMeCgluAla----- 112
DB 880 GCCAACAAGAAAGAGGTTCAATGACTCCGGAAGCAGTGTCTGTGAATAAGGTA 821
QY 113 VallysGlnUlyLeuValGluPhetPalaYsGlnAla-----GluGlu 129
DB 820 AACTGAAAGATGCTCTTCTTTTGTGAACCGGCTTCAAGCTTAAAGCAAGAAAG 761
QY 130 VallysValGlnUlyLeuValGluMetGlnApsrThrAsnAlaLeuYsAps 149
DB 760 ACCTCAAGCCCTGTGTTCCAGAAAG-----NATTT-CGG 726
QY 150 GlnLeuPheValGlnUlyLeuAlaArgHisIleLeuValYsThrGluAps-- 168
DB 725 CAACCTMTTCCGAACAAGCGGATTCGCCCAATCTCTATGAGGTGAACGACAG 666
QY 169 -----GluAlaYsArgIleIleSerGluIleApsYsGlnProYsAlaYs 184
DB 665 GTCCGCGACGACGAGCCAGCGCAAGATCGACGAGATC-----AAGGCTCGCTGGCC 612
QY 185 LysGlnUlyLeuPheIleGlnUlyLeuAlaAsnArgAspThrIleApsProAsnSerYsAps 204
DB 611 AAGGCGAGAGATTTCCGCGCTGCGCAAGATTCTCCAGATATCCGCTCG----- 558
QY 205 AAlaGlnAnGlyYsApsLeuGlyYsPheGlnUlysApsGlnMetAlaProApsPheSer 224
DB 557 GCCGCGACCGCGGCTGAGCTGCGCTACGCGGCTCCGCGCTAGACACCCCGCTTCAG 498
QY 225 LysAlaAlaPheAlaLeuThrProGlyApsYrThrYsThrProValYsThrGluPhe 244
DB 497 GAGGCGCTGTATGCGCTGAAAGCAAGGTAG--GTAATCGCCCGGTGAAGACTCCGTAAC 441
QY 245 GlyYrThrIleIleYrThrLeuIleSerYsApsSerProValThrYrThrYrGluGln 264
DB 440 GCGTCAACCTGATCAAGCTGTGGCGCTGACGAGCGCGCAAGTACCGGAGCTGGAAGAC 381
QY 265 AAlaYsProThrIleLeuGlyMetLeuGlnUlyLeuPheGlnUlyArgMetApsGln 284
DB 380 CTCAGACCGAAGCTCGAGAGCACTGAAGAAACAGATGCTGACAGCGCTTCGTCAAG 321
QY 285 ArgIleGlnUlyLeuArgYsHisAla 293
DB 320 GCTACCAAGAGCTGGAAGCTCCGCC 294

RESULT 14

BZ574276

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

Location/Qualifiers
1..1540
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msb2_3590"
/clone_1ib="msb"
/note="Environmental isolate. Whole genomic shotgun library."

FEATURES

Alignment Scores:
Pred. No.: 2,496-06 Length: 1540
Score: 152.50 Matches: 41
Percent Similarity: 52.26% Conservative: 57
Best Local Similarity: 26.45% Mismatches: 17
Query Match: 10.03% Indels: 17
DB: 8 Gaps: 6

ORIGIN

US-10-039-183a-2 (1-299) x BZ574276 (1-1540)
QY 153 VallysGlnUlyLeuAlaArgHisIleLeuValYs-----Thrglu 167
DB 206 GTCCGTGACGAAGTCAATGCTCCGCAATCTCTGTCACAGCCCAAGCAATCCGACGAA 265
QY 168 ApsGlnUlysArgIleIleSerGluIleApsYsGlnProYsAlaYsGlnAla 187
DB 266 GCGGAGACCGAAGACTGCGCGCAAGCTGTGACGCGCATCAGTCCGCGGAGAGAC-- 322
QY 188 LysPheIleGlnUlyLeuAlaAsnArgAspThrIleApsProAsnSerYsApsAlaGln 207
DB 323 ---TTCCGCGAAGCTGCGCAAGCTTCCGCAAGATCCGGGTTCC-----GCCTCAAC 373
QY 208 GlyYsApsLeuGlyYsPheGlnUlysApsGlnMetAlaProApsPheSerYsAlaAla 227
DB 374 GCGGCGACCTGAACTGATGATCGAAGCCCTGCGCGCAATTCCTGACGTGATG 433
QY 228 PheAlaLeuThrProGlyApsYrThrYsThrProValYsThrGluPheGlyYrHis 247
DB 434 AACGATACCGCGAGGCGAGGCTGCAAG--CCGTTCCGCTCCAGTTGCGTGGCAGC 490
QY 248 IleIleYrLeuIleSerYs-----ApsSerProValThrYrThrYrGluGln 264
DB 491 ATCCGACAGTCTTCCGCGCTGCGCCACCGACGACGAGAACTTCGCGAGACGAG 550
QY 265 AAlaYsProThrIleLeuGlyMetLeuGlnUlyLeuPheGlnUlyArgMetApsGln 284

Db 551 GCG-----GTAAGCGTACTGCGCGAACCGCAAGTACGACGAGGAACTCGCAAGCC 598
Qy 285 ArgIleGluGluLeuArgIleGlyHisIleValIleAsnLys 299
Db 599 TGGCTGCGCGACAGTCCGCGACGAGGCTTACGTGGAATCAGCAG 643

RESULT 15
CF843490/c
LOCUS 502 bp mRNA linear EST 30-OCT-2003
DEFINITION pSHB025XG18f USDA-IFARS:Expression of Phytophthora sojae genes during infection and propagation_sHB Phytophthora sojae cDNA clone sHB025G18 5, mRNA sequence.

ACCESSION CF843490
VERSION CF843490.1 GI:38059144
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
1 (bases 1 to 502)
REFERENCE Tyler,B. Not Published
AUTHORS Tyler,B. Not Published
JOURNAL Unpublished (2003)
COMMENT Contact: Tyler B
Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmyler@vt.edu
PCR Primers
FORWARD: BK reverse primer
BACKWARD: BK reverse primer
Plate: 025 row: G column: 18
Seq primer: BK reverse primer
High quality sequence stop: 502.
Location/Qualifiers

FEATURES

source
1..502
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="SHB025G18"
/tissue_type="mycelium"
/cell_line="P6497"
/dev_stage="48 hr. post infection stage"
/lab_host="Soybean plant"
/clone_lib="USDA-IFARS:Expression of Phytophthora sojae genes during infection and propagation_sHB"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Alignment Scores:
Pred. No.: 7.16e-07 Length: 502
Score: 151.50 Matches: 36
Percent Similarity: 54.95% Conservative: 14
Best Local Similarity: 39.56% Mismatches: 34
Query Match: 9.97% Indels: 7
DB: 7 Gaps: 3

US-10-039-183a-2 (1-299) x CF843490 (1-502)

Qy 164 VallysThrgLAspGluAlaLysArgIleIleSerGluIleAspLysGlnProLysAla 183
Db 500 GTGCCGACCGAGGAGGAGCGACGACGATCTTGAAGAGCTG-----CAGCGGCGCGAC 447
Qy 184 LysLysGluAlaLysPheIleGluLeuAlaAsnArgAspThrIleAspProAsnSerLys 203
Db 446 GACCTGAGGCGCACGCTTGGCGGCTGGCGAAGAGCGGACGACGATGTCCTCACTGC 387
Qy 204 AsnIleGlnAsnGlyGlyAspLeuGlyLysPheGlnLysAsnGlnMetAlaProAspPhe 223
Db 386 -----MAGGAGGAGGAGCTTGGGCTCTTGGTCGCGGCGACGATGTCCTCCAGTTTC 336

Qy 224 SerLysAlaAlaPheAlaLeuThrProGlyAspTyrThrLysThrProValLysThrgL 243
Db 335 GACAAAGTGGGCTTGAGAAAGCCCGTGGGGAGCTCCACAAG-----GTCAAGACGCG 282
Qy 244 PheGlyTyrHisIleIleIleTyrLeuIleSerLys 254
Db 281 TTTGATGGCACTTGTCGTGATCAAGCGCGC 249

Search completed: April 24, 2005, 11:19:59
UOD time : 3362.11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2005, 02:20:20 ; Search time 724.259 Seconds

(without alignments)
3261.231 Million cell updates/sec

Title: US-10-039-183A-4

Perfect score: 2043
Sequence: 1 MAKEFNRTKPHVNIIGTIGH.....FAIRGRTYGVAGVSNIIIE 399

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=x1h
-Q=/cgn2.1/USPTO/US10039183/runat_22042005_122424_20098/app_query.fasta_1.1038
-DB=N GeneSeq libDec04 -QFMT=fasta -SUFFIX=ring -MINMATCH=0.1 -LOOFCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=humand0.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=prco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10039183_@CGN_1_1_796.@runat_22042005_122424_20098 -NCPU=6 -ICPU=3
-NO MMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_GeneSeq_16Dec04:*
1: geneSeqn1980s:*
2: geneSeqn1990s:*
3: geneSeqn2000s:*
4: geneSeqn2001as:*
5: geneSeqn2001bs:*
6: geneSeqn2002as:*
7: geneSeqn2002bs:*
8: geneSeqn2003as:*
9: geneSeqn2003bs:*
10: geneSeqn2003cs:*
11: geneSeqn2003ds:*
12: geneSeqn2004as:*
13: geneSeqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2043	100.0	1200	4	AAS53681 Helicobac
2	2043	100.0	1200	8	ACA34860 Prokaryot
3	2043	100.0	1448	2	AAV07964 Helicobac
4	1772	86.7	1200	8	ACA30274 Prokaryot
5	1609.5	78.8	17782	3	AAA81530 N. mening

6	1609.5	78.8	110000	3	AAA81490_01	Continuation (2 of
7	1609.5	78.8	349980	4	AAFI21544	AAFI21544 Nelsieria
8	1608	78.7	1185	3	AAS51250	AAS51250 Enterococ
9	1608	78.7	1185	8	ACA18405	ACA18405 Prokaryot
10	1608	78.7	1185	12	ADO25365	ADO25365 E.faecali
11	1608	78.7	1185	4	AAS52748	AAS52748 Enterococ
12	1606.5	78.6	1185	8	ACA41574	ACA41574 Prokaryot
13	1606.5	78.6	1185	8	ACA41586	ACA41586 Prokaryot
14	1606.5	78.6	65632	3	AAA81502	AAA81502 N. mening
15	1600	78.3	1185	12	ADO25367	ADO25367 E.faecali
16	1598.5	78.2	1182	8	ACA41360	ACA41360 Prokaryot
17	1598.5	78.2	1182	10	ABZ41688	ABZ41688 N. gonorr
18	1598.5	78.2	1182	10	ABZ41220	ABZ41220 N. gonorr
19	1597.5	78.2	1188	8	ACA26430	ACA26430 Prokaryot
20	1592.5	77.9	110000	6	ABG67196_4	ABG67196_4
21	1592	77.9	110000	6	ABG69245_28	ABG69245_28
22	1592	77.9	110000	6	ACA36873	ACA36873 Prokaryot
23	1586	77.6	1188	8	ABA03041_27	ABA03041_27
24	1586	77.6	110000	6	ACA41337	ACA41337 Prokaryot
25	1585.5	77.6	1186	8	ACA26743	ACA26743 Prokaryot
26	1581.5	77.4	1188	8	ACA26802	ACA26802 Prokaryot
27	1581.5	77.4	106645	13	ADT05645	ADT05645 Haemophil
28	1579.5	77.3	1185	4	AAS53325	AAS53325 Haemophil
29	1579.5	77.3	1185	4	AAS53335	AAS53335 Haemophil
30	1579.5	77.3	1185	8	ACA34171	ACA34171 Prokaryot
31	1579.5	77.3	1185	8	ACA34190	ACA34190 Prokaryot
32	1579.5	77.3	25523	13	ADT05538	ADT05538 Haemophil
33	1579.5	77.3	106645	13	ADT05645	ADT05645 Haemophil
34	1579.5	77.3	110000	2	AAT42063_05	AAT42063_05
35	1579.5	77.3	110000	2	AAT42063_06	AAT42063_06
36	1579.5	77.3	1182	8	ACA47619	ACA47619 Prokaryot
37	1578.5	77.2	1188	8	ACA25058	ACA25058 Prokaryot
38	1576.5	77.0	1182	8	ACA46415	ACA46415 Prokaryot
39	1572.5	77.0	1185	8	ACA43200	ACA43200 Prokaryot
40	1572.5	77.0	1245	6	ABN93077	ABN93077 Staphyloc
41	1572.5	77.0	1245	13	ADS02331	ADS02331 Staphyloc
42	1572.5	77.0	1245	6	AAI12938	AAI12938 Enterococ
43	1572	76.9	4315	2	ABN98733	ABN98733 Enterococ
44	1572	76.9	4315	6	AAA51937	AAA51937 Staphyloc
45	1571.5	76.9	1182	4	AAA51937	AAA51937 Staphyloc

ALIGNMENTS

RESULT 1	AAS53681	standard; DNA; 1200 BP.
ID	AAS53681	
XX		
AC	AAS53681;	
XX		
DT	13-FEB-2002 (first entry)	
XX		
DE	Helicobacter pylori DNA for cellular proliferation protein #135.	
XX		
KW	Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;	
XX	antibacterial; drug design.	
KW	antibacterial; drug design.	
XX		
OS	Helicobacter pylori.	
XX		
PN	WO200170955-A2.	
XX		
PD	27-SEP-2001.	
XX		
XX		
PF	21-MAR-2001; 2001WO-US009180.	
XX		
PR	21-MAR-2000; 2000US-0191078P.	
XX		
PR	23-MAY-2000; 2000US-0206848P.	
XX		
PR	26-MAY-2000; 2000US-0207727P.	
XX		
PR	23-OCT-2000; 2000US-0242578P.	
XX		
PR	27-NOV-2000; 2000US-0253625P.	
XX		
PR	22-DEC-2000; 2000US-0257931P.	
XX		
PR	16-FEB-2001; 2001US-0269308P.	
XX		

PA (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
DR P-PSDB; AAU35822.
XX
PT New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 7318; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the genes,
their use in the discovery of novel antibiotics, the essential genes
themselves and the encoded proteins. The prokaryotes used are *Escherichia*
coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
useful for the identification of potential new targets for antibiotic
development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
antibodies capable of binding to the expressed proteins. The proteins can
be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

Sequence 1200 BP; 365 A; 207 C; 319 G; 309 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,47e-191 Length: 1200
Score: 2043.00 Matches: 399
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-039-183a-4 (1-399) x AAS53681 (1-1200)

QY 1 MetAlaIySGluYpheaSnaRgThRlySPRoH:isValAsnIleGlyThrIleGlyHis 20
DB 1 ATGGCAAAAGAAAGATTAAACAGAACTAAGCCGCACTGTAATATTGGAACCAATTGGGCAT 60
QY 21 ValAspHisGlyLeuThRleuSerAlaAlaIleSerAlaValLeuSerLeuysGly 40
DB 61 GTAGACCAATGGTAAACGACTTTGAGTGCAGCGATTTTCAGCGGTGCTTTCTTTGAAAGGT 120
QY 41 LeuAlaGluMetLlyAspTyraSPaenIleAspAsnAlaPProGluGluysGluArgGly 60
DB 121 CTTGCAGAATAAGAAAGACTATGATATATGATTAACGCCCTCGAAGAAAGAAAGAGAGG 180
QY 61 IletPtlleAlaThSerHisIleGlyuTyrgluThGluAsnArgHisTyraHisVal 80
DB 181 ATCACTATGCTACTCTTCACATGTAATAGAACTGAAACAGACACTATGCGCATGTG 240
QY 81 AspCySPProGlyHisIleAspTyraVallyAsnMetIleThGlyAlaAlaGluMetAsp 100
DB 241 GATTGCCCGACGACGCTGACTATGTAAACATGATCAACCGGTGCGCGCAATGAGC 300
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyPProMetProGluThRArgGluHisIle 120
DB 301 GGAGCGATTTGGTGTCTGCTGACGCTAGGCCCTTACTCAAACTAGGAGCATATC 360
QY 121 LeuLeuSerArgGluValGlyValProHisIleValValPheLeuAsnLysGluAspMet 140
DB 361 TTATTGTCTCGTCAAGTAGCGGTGCTCACTCGTTGTTCTTTAAACAAAGACAG 420
QY 141 ValAspAspGluGluLeuGluValGluMetGluValArgGluLeuLeuSerAla 160

DB 421 GTAGATGACCAAGATTTGTAGAACCTTGAGAAATGGAATGCGCAATTTGTGACCGCG 480
QY 161 TyrgluPheProGlyAspAspThRProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 481 TATGAAATTTCTGGGATGACACTCTCTATCTGAGGGGTTGACGCTTTAAGACCTTTAGAA 540
QY 181 GluAlaIySAlaGlyAsnValGlyuTyrglyGlyLysValLeuLysLeuMetAlaGlu 200
DB 541 GAGCAAGAGCTGGTGAATGTGGGTGAATGGGTGAAAAAGTCTTAACTTATGGCTGA 600
QY 201 ValAspAlaTyrlleProThRProGluArgAspThRgluYpheaSPRoVal 220
DB 601 GTGATGCTTATATCTTACTTACCAAGAAAGACACTGAAAAAACTTTCTTGATCGCGTT 660
QY 221 GluAspValPheSerIleAlaGlyArgGlyThRValValThGlyArgIleGluArgGly 240
DB 661 GAAAGATGTTCTTCAATGCGGGTAGAGGACTGTGGTTACAGGTGAGATTGAAAGAGGC 720
QY 241 ValValIySValGlyAspGluValGluIleValGlyIleArgPProThRgluYpheaSPR 260
DB 721 GTGGTGAAGTGAAGCGATGAAGTGAATCGTTGGTATCAACCTACACAAAGAAAGCACT 780
QY 261 ValThRglYValGluMetPheaRglYSGluLeuGluLysGlyGluAlaGlyAspAsnVal 280
DB 781 GTAACCGGTGTGAATGTTTAGAAAGAGTTGAAAGAAAGGTAAGCGCGCATATATGTG 840
QY 281 GlyValIleLeuAsnArgGlyThRlySylSGluGluValGluArgGlyMetValLeuLys 300
DB 841 GCGGTGCTTTGAGAGGAATGTTAGAAAGAGTTGAAAGAAAGCGGTATGTTCTATGCAAA 900
QY 301 ProGlySerIleThRProHislySPheGluGlyGluIleTyraValLeuSerLysGlu 320
DB 901 CCGGTTCTATCACTCCGCAAGAAATTTAGAGGAGAAATTTATGCTCTTTAAAGAA 960
QY 321 GluIyGlyArgHisThRProPheThRAsnTyraSPProGluPheTyraValArgThr 340
DB 961 GAAGCGGAGACACACTCCATTTCTTACCAATTAACCGCCCAATTTCTATGTGCGCACA 1020
QY 341 ThrAspValThGlySerIleThRleuProGluGlyValGluMetValMetProGlyAsp 360
DB 1021 ACTGATGACCTGCTCTATACCCCTTCTGAAGCGTGAAGATGTTATGCTGCGCAT 1080
QY 361 AsnValIySylleThRValGluLeuLysSerProValAlaLeuGluLeuGlyThRlySPhe 380
DB 1081 AATGTGAATATCACTGATGATTTAGCCCTGTGCGTTAGAGTTGGGAATTAATTT 1140
QY 381 AlaIleArgGluGlyArgThRValGlyAlaGlyValIleSerAsnIleIleGlu 399
DB 1141 GCGATTCGTGAAGCGGTAGAACCTGTGTGCTGTGTTGTGACCAATTTATTTGAA 1197
RESULT 2
ACA34860
ID ACA34860 standard; DNA; 1200 BP.
AC ACA34860;
XX
XX 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #16517.
DE
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
XX Helicobacter pylori.
OS
XX
XX WO200271183-A2.
PN
XX
PD 03-OCT-2002.
PF
XX 21-MAR-2002; 2002WO-US009107.
PF
XX 21-MAR-2001; 2001US-00815242.
PR

PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zvekind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.
DR P-PSDB; ABU30990.

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 22730; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1200 BP; 365 A; 207 C; 319 G; 309 T; 0 U; 0 Other;

Alignment Scores:

Score: 2,476-191 Length: 1200
Percent Similarity: 2043.00 Matches: 399
Best Local Similarity: 100.00% Conservative: 0
Query Match: 100.00% Mismatches: 0
DB: Indels: 0
Gaps: 0

US-10-039-183a-4 (1-399) x ACA34860 (1-1200)

QY 1 MetAlaValGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
DB 1 ATGGCAAAAGAAAGATTAACTAAAGCCGCACTGTAATATTTGGAACCAATGGGCAT 60
QY 21 ValAspHisGlyLysThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuGly 40
DB 61 GTAGACCAATGTAAGCAAGCTTTGAGTGACGCAATTCACGCGTGTCTTTGAAAGGT 120
QY 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60

DB 121 CTTCGCAAAATGMAAGACTATGATATATGTATACGCCCTGAGAAAAAGAAAGAGG 180
QY 1 IeThrIleAlaThrSerHisIleGlyLysThrGluAsnArgHisIleTyrAlaHisVal 80
DB 61 ATGCTATGCTACTCTTCTCACTGAAATGTGACCTGAAAACAGACACTATGCGCATGTG 240
QY 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGlnMetAsp 100
DB 241 GATTCGCCAGACACGCTGACTGACTATGTAATAAACAATGATCACCGGTGCGCAATGAC 300
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
DB 301 GGAGCGAATTTGGTTGTTCTGACAGCTATGCGCCCTTAATGCTCAAGAGCAATATAC 360
QY 121 LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMet 140
DB 361 TTATGTCTCGCAAGTAGGCGTGTCTCACTGCTGTTGTTTAAACAAACAGACATG 420
QY 141 ValAspAspGlnGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
DB 421 GTAGATGACCAAGATTTGTAGAACTTGTAGAAATGGAAGTCCGCAATTTGTAGCGCG 480
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 481 TATGAATTTCTGCGGATGACACTCTCTATCGTAGCGGTTTCACTTAAAGACTTTAGAA 540
QY 181 GluAlaValAlaGlyAsnValGlyGluThrProGlyLysValLeuLysLeuMetAlaGlu 200
DB 541 GAAGCAAAAGCGCTGTAATGTGGTAAATGGGTGAAAGTCTTAATATTAATGCTGAA 600
QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
DB 601 GTGATGCTATATCTCTACTCTCAAGAAAGACATGAAAAAATCTTCTGATGCCGTT 660
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
DB 661 GAAGATGTCTCTATTTGCGGTAGAGGACCTGTGTACAGTAGTAGTTGAAAGAGCG 720
QY 241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThr 260
DB 721 GTGTGAAAGTAGGGAATGGAATCGTGTATACAGCTACACAAAAACGACT 780
QY 261 ValThrGlyValAlaGluMetPheArgLysGluLeuGluLysGlyValAlaGlyAspAsnVal 280
DB 781 GTACCGGTGTGAAATGTTTGTGAAAGATTTGAAAAAGTGAAGCCGCAATATGTG 840
QY 281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys 300
DB 841 GCGGTGCTTTGAGAGGAACCTAAAAAGAAAGTGAACGCGGTATGTTCTATGCAAA 900
QY 301 ProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGlu 320
DB 901 CCAGGTTCTATCACTCCCAACAAAGAAATTTGAGGAATTAATGCTCTTTCTAAAGAA 960
QY 321 GluGlyLysArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArgThr 340
DB 961 GAAGCGGAGACACACTCCATCTTCTACCAATTAACGCCCAATCTATATGCGCACA 1020
QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
DB 1021 ACTGATGTGACTGCTCTATCACTCTCTCTGAAAGCGTAGAAATGTTATGCTGCGCAT 1080
QY 361 AsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrLysPhe 380
DB 1081 AATGTGAAATCACTGTAGAGTGTATGCGCTGTGCGTGTAGATTTGGAACCTAAATTT 1140
QY 381 AlaIleArgGluGlyLysThrValGlyAlaGlyValAlaSerAsnIleIleGlu 399
DB 1141 GCGATTCGTGAAGCGGTAGACCTGTGCTGTGTGTGTGACGAAATTTATTGA 1197

RESULT 3
AAV07964
ID AAV07964 standard; DNA; 1448 BP.

XX AAV07964;
 AC
 XX 25-MAR-2003 (revised)
 DT 02-FEB-1999 (first entry)
 XX Helicobacter pylori 50 kDa polypeptide GHPO 750 DNA.
 DE
 XX GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; ss.
 KM
 XX Helicobacter pylori.
 OS
 XX Key Location/Qualifiers
 FH 118.1317
 FT CDS /*tag= a
 FT
 XX MO9843479-A1.
 PN
 XX 08-OCT-1998.
 PD
 XX 31-MAR-1998; 98MO-US006421.
 PF
 XX 01-APR-1997; 97US-00831310.
 PR 01-APR-1997; 97US-00834666.
 XX
 PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PA (HUMA-) HUMANA GENOME SCI INC.
 XX
 PI Kleantous H, Lissolo L, Tomb J, Miller C, Algarawi A;
 DR WPI; 1998-568251/48.
 DR P-PSDB; AAW73035.
 XX
 PT New isolated Helicobacter polynucleotides - used to develop products for
 PT the diagnosis, prevention and treatment of Helicobacter infections and
 PT gastroduodenal diseases.
 XX
 PS Claim 1; Page 150-152; 184pp; English.
 XX
 CC This DNA sequence codes for a 50 kDa Helicobacter pylori polypeptide (see
 CC AAW73035) designated GHPO 750. A polynucleotide encoding GHPO 750 can be
 CC obtained from genomic DNA by PCR amplification (see AAV07971-72. The
 CC invention provides polynucleotides (see AAV72001, AAV07912-21 and
 CC AAV07963-64) encoding a family 76 kDa Helicobacter polypeptides (see
 CC AAW73022-32), GHPO 750 and a 32 kDa polypeptide (see AAW73034). These
 CC polynucleotides were initially identified in a search of H. pylori
 CC genomic databases. DNA cassettes for expression of the Helicobacter
 CC proteins (unprocessed or mature forms) in prokaryotic or eukaryotic cells
 CC are provided. The polynucleotides can be used in vaccines to prevent or
 CC treat Hb infection in a mammal. Viral (especially poxvirus) or bacterial
 CC vectors are used. Products and methods of the invention allow treatment
 CC and prevention of gastroduodenal diseases associated with Hb infections,
 CC including acute, chronic, and atrophic gastritis, and peptic ulcer
 CC diseases, e.g. gastric and duodenal ulcers. Diagnostic and detection
 CC methods are also provided. GHPO 750 was demonstrated to be a protective
 CC antigen. (Updated on 25-MAR-2003 to correct PI field.)
 CC
 XX Sequence 1448 BP; 466 A; 237 C; 364 G; 381 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 3,16e-191 Length: 1448
 Score: 2043.00 Matches: 399
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-039-183a-4 (1-399) x AAV07964 (1-1448)

QY 1 MetAlaIyGluIyPheAsnArgThyIyPProHisValAsnIleGlyThrIleGlyHis 20
 DB 118 ATGGCAAAAGAAAGATTAAACAGAACTAAGCCGCACTGTTAATATGGAAACCAATGGGCAT 177

QY 21 ValAspHisGlyIySerThrThiLeuSerAlaAlaIleSerAlaValIleSerIleuIyGly 40
 DB 178 GTAGACCATGTAAGAAACGACTTGTAGTCAGCGATTTTCAGCGTCTTCTTGAAGGT 237
 QY 41 LeuAlaGluMetIyAspTyrAspAsnIleAspAsnAlaProGluGluIySerIyGly 60
 DB 238 CTTCGAGAAATGAAAGACTATGATATATGTAAACGCCCTGAAGAAAGAAAGAGGG 297
 QY 61 IleThrIleAlaThrSerHisIleGlyIySerGluThrGluAsnArgHisTyrAlaHisVal 80
 DB 298 ATCACTATCGCTACTTCTTCACATTGAATATGAGATCGAAAACAGACATATGCGCATGG 357
 QY 81 AspCysProGlyHisAlaAspTyrValIyAsnMetIleThrGlyAlaAlaGluMetAsp 100
 DB 358 GATGCCCAAGAGCAGCGTCACTATGTAAACATGATACACCGGTGGCGCAATATGAC 417
 QY 101 GlyAlaIleLeuValIleSerAlaAlaAspGlyProMetProGluThrArgIleHisIle 120
 DB 418 GGAGCGAATTTGGTGTGTTCTGACAGCTGATGGCCCTATGCTCAAACTAGGAGCATATTC 477
 QY 121 LeuIleSerArgGluValIleGlyValIleProHisIleValIlePheLeuAsnIySerIleAspMet 140
 DB 478 TTATTGCTCTGTCAGATGAGCGGTGCTCATCATGCTGTTCTTTCTTAAACAAACAGACATG 537
 QY 141 ValAspAspGluGluLeuLeuGluIleuValGluMetGluValArgGluIleuLeuSerAla 160
 DB 538 GTAGATGACCAAGAAATGTTAGAACTTGTAGAAATGAAATGAAATGCGGAATGTTGAGCGCG 597
 QY 161 TyrGluIleProGlyIyAspAspThrProIleValAlaIleGlySerAlaIleuArgAlaIleuGlu 180
 DB 598 TATGAATTTCCGCGGATGACACTCTCTATCGTAGCGGTTCAGCTTTAAGACCTTTAGAA 657
 QY 181 GluAlaIyAlaGluValAsnValIleGlyIyIleProGlyIyGluIyValIleuIySerIleuMetAlaGlu 200
 DB 658 GAAGCAAGCGCTGTAATGTGGGTAAATGGGTGTAAGGCTTAACTTATGCTGAA 717
 QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluIySerIlePheIleuMetProVal 220
 DB 718 GTGATGCTATATCTTATCTTACCTTACCTCAGAAAGACATGAAACCTTCTGATGCGGTT 777
 QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValIleValIleGluArgIleGluArgGly 240
 DB 778 GAAGATGCTTCTTATTTGCGGTGAGAGGATGTTGTTACGTTACGTTAAGAAAGAGCG 837
 QY 241 ValIleValIyAlaGluAspGluValGluIleValIleGlyIleArgProThrGluIySerThrThr 260
 DB 838 GTGGTGAAGATGAGCGATGAAAGTGAATCGTTGTATCAAGCTTACACAAACAAACGACT 897
 QY 261 ValThrGlyValGluMetPheArgIySerGluIleuGluIyGlyGluAlaGlyAspAsnVal 280
 DB 898 GTTAACCGGTGTAAGAAATGTTTGAAGAGGTTGAAAGAAAGGTGAAGCCGCGCATATGTTG 957
 QY 281 GlyValIleLeuAsnArgGlyThrIyIySerGluIyValIleGluArgGlyMetValIleuIySerIy 300
 DB 958 GGCGGTCTTTGAGAGGAACTTAAAGAAAGAGGAGAACCGGTATGTTGTTTATGCAAA 1017
 QY 301 ProGlySerIleThrProHisIyIySerPheGluIyGluIleThrValIleuSerIyGlu 320
 DB 1018 CCAAGTTCATACCTCCGACAAAGAAATTTGAGGAGAAATTTATGCTTCTTAAAGAA 1077
 QY 321 GluIyGlyArgHisIleThrProPhePheThrAsnTyrArgProGluPheTyrValArgThr 340
 DB 1078 GAAGCGGGAGACACATCCATTTCTTCAACCAATTAACCGCGCAATTTCTATGTGGCA 1137
 QY 341 ThrAspValThrGlySerIleThrIleuProGluIyValIleGluMetValIleMetProGlyAsp 360
 DB 1138 ACTGATGTGACTGCTCTATCAACCTTCTCTGAGGCGTGAAGATGTTATGCTGCGCAT 1197
 QY 361 AsnValIySerIleThrValIleuIleSerProValAlaIleuGluIleuGlyThrIyPhe 380
 DB 1198 AATGTGAAATACATGTAAGATGATTAAGCCCTGTGGTGAAGTGGAACTTAATTT 1257
 QY 381 AlaIleArgGluIyGlyArgThrValIleGlyAlaIyValIleSerAsnIleIleGlu 399

Db 1258 GCGATTCGTGAAGCGGTAGACCGTGTGCTGTGTTGTGAGCAATATTATTA 1314

RESULT 4
ID ACA30274 standard; DNA; 1200 BP.
XX ACA30274;
XX ACA30274;
XX 19-JUN-2003 (first entry)
XX prokaryotic essential gene #11931.
XX
XX Antisense; deg. prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Campylobacter jejuni.
XX MO20027183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forbych RA, Xu HH;
XX MPI; 2003-029926/02.
XX P-PSDB; ABU26404.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 18144; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcr_sequences
XX
SQ Sequence 1200 BP; 367 A; 185 C; 255 G; 373 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,196-164 Length: 1200
Score: 1772.00 Matches: 336
Percent Similarity: 93.23% Conservative: 36
Best Local Similarity: 84.21% Mismatches: 27
Query Match: 86.74% Indels: 0
DB: 8 Gaps: 0

US-10-039-183A-4 (1-399) x ACA30274 (1-1200)

QY 1 MetAlaLVGLuVpPheAsnArGThrLyPProHISValaenIleGIYThrIIeGIYHIS 20
DB 1 ATGGCTAAAGAAATTTTCAAGTAAATAGCCACGTAATATTTGTAATTTGCTCAT 60
QY ValaAspHISGIYLYeThrThrIleuSerAlaAlaIleSerAlaValIleuSerIleuVSGIY 40
DB 61 GTTGACCATGGTAACCTACTTACACGCTGCTATTCTGCTGTTCTTTCTTGAAAGAGT 120
QY 41 LeuAlaGIuMeTLYaSPTYrAspAsnIleAspAsnAlaProGIuGIuVSGIuArgGIY 60
DB 121 TTAGCAGAGCTTAAAGATTATGATATATCATATATGCTCCAGAAAGAAAGAGCOTGGT 180
QY 61 ILeThrIIeAlaThrSerHISIIeGIuTYrGIuThrGIuAsnArGHSIYrAlaHISVal 80
DB 181 ATTACTATTTGCTACTTCTCATATTTGATATGTAAGAACGCAATCGTCACATCTACACGCTT 240
QY 81 AspCySPROGIIYHISAlaSPTYrValIYAsnMeTIIeThrGIYAlaIleGIuMeTAsp 100
DB 241 GACGCCCCAGGTCAGCAGATTAATATGTAATTAACATGATTAACAGGCTGCACAAATGAT 300
QY 101 GIYAlaIleIleuValIleSerAlaAlaAspGIYProMeTProGIuThrArgGIuHISIle 120
DB 301 GAGGCGATCTTGCTGTTGTTCTGCTGCACATGCGCTTAATGCCAATAGAGCAGCAT 360
QY 121 LeuIleuSerArGIuValIleGIYAlaProHISIIeValIlePheIleuAsnLYeGIuAspMeT 140
DB 361 CTTCTTCTCGTCAAGTACGCTTCATATATTTGTTTATATATTAACAGATTAAG 420
QY 141 ValaAspAspGIuGIuIleuGIuIleuValIleGIuMeTGIuValIleArgGIuIleuIleuSerAla 160
DB 421 GTTGATGATGCTGAACCTTTAGAGTATGTAAGAAATGAAATTAAGAAATTAAGCTT 480
QY 161 TYrGIuPheProGIYAspAspThrProIleValaIleGIYSerAlaIleuArgAlaIleuGIu 180
DB 481 TATGATTTCCAGGCGATGATACACTATTTATTTGCTGCTGCTTAAACCTCTTGAA 540
QY 181 GIuAlaIleValaGIYAsnValaGIYGIuTYrGIYGIuIleValIleuIleuMeTAlaGIu 200
DB 541 GAACTAAAGCTGCAAGATGATGTAATGCTCAGCAAAATATATGATCTTTATGCTGCA 600
QY 201 ValaAspAlaTYrIIeProThrProGIuArgAspThrGIuIleuThrPheIleuMeTProVal 220
DB 601 GTTGATACCTATATTCACACTCAACTGCTGATCTGTAAGAAAGAAAGCTCTTGATGCCAAT 660
QY 221 GIuAspValIleSerIIeAlaGIYArgGIYThrValIleThrGIYArgIIeGIuArgGIY 240
DB 661 GAAAGTCTTTCTCAATTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 241 ValValIleValaGIYAspGIuValaIleuIleValaGIYIIeArgProThrGIuIleuThr 260
DB 721 GTTGTAAGAAAGTGAATCAATCAAAATCGTGTATTAATTAAGAAATCAAACTCACT 780
QY 261 ValThrGIYValaGIuMeTPhaArgIleGIuIleuGIuIleuValaGIYAspAsnVal 280
DB 781 GTAACAGGTGTGAATGTGTAAGAAAGAAATGATCAAGGGAAGGAGATTAACGTA 840
QY 281 GIYValIleuIleuArgGIYThrIleYleGIuGIuValaGIYArgGIYMeTValIleuCyIleu 300

841 GGTGTTCTTCTTCTGCTGACTATAAAAAAGAGAGTATCCGTGTAAGTTCCTGCTAA 900
301 ProglSerIleThProHisIleLysPhegluIleTyValIleuSerIleGlu 320
901 CCAATATCAATTACTCCACACCTGACTTCGAACTGAAAGTTATATCTTAATAATGAAT 960
321 GluIglYIaYrGHIeThProPhePheThAsnTYrArgProGluPheTYrValIaYrThr 340
961 GAAGGTGTAGACATACCTCATCTTTTAAACAATAAGCACACAGTTTATATGAAGACA 1020
341 ThrAspValThrGlySerIleThIleuPProgluIglValGluMetValMetProglYasp 360
1021 ACTGATGTATACGGTTCATTAATAATGCTGATGGTGAATGGTATATGCCAGCTGA 1080
361 AsnValYIleThValIgluIleuSerProValAlaIleuGluIleGlyThrIleYspHe 380
1081 AATGTAGAAATTAATCTGTAAGCTTGAATCGCTCAGTAGCACTTGAAAGAAAGAACTGTTT 1140
381 AlaIleArgIgluIglYIaYrGThrValIglYalIaYIaValSerAsnIleIleGlu 399
1141 GCTATTCGTAAGGTGTAAACCTGTGGTTCAGGCTGTTGTTCTTAATAATTATTA 1197

RESULT 5

AAAB1530
ID AAAB1530 standard; DNA; 17782 BP.

AAAB1530;

04-DEC-2000 (first entry)

N. meningitidis partial DNA sequence gnm_77 SEQ ID NO:77.

Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
antigen; vaccine; diagnosis; infection; antibacterial; identification;
Meningococcus B; Memb; ds.

Neisseria meningitidis.

MO200022430-AZ.

20-APR-2000.

08-OCT-1999; 99WO-US023573.

09-OCT-1998; 98US-0103794P.

30-APR-1999; 99US-0132068P.

(CHIR) CHIRON CORP.

Frazer CM, Hickey E, Peterson J, Tettein H, Venter JC;
Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
Rappuoli R, Pizza M;

WPI; 2000-318079/27.

Isolated nucleotide sequences of Neisseria meningitidis which can be used
in the diagnosis and treatment of N. meningitidis infection and other
Neisserial infections, for example, N.gonorrhoea.

Claim 7; Page 1457-1462; 1760DP; English.

The present invention describes methods of obtaining immunogenic proteins
from Neisseria genomic sequences. AAAB1453 to AAAB2414 represent
specifically claimed Neisseria meningitidis genomic DNA sequences;
AAAB1260 to AAAB1303 and AAAB25620 to AAAB25663 represent Neisseria DNA
sequences and their corresponding proteins; AAAB1254 to AAAB1259 and
AAAB1304 to AAAB1321 represent PCR primers used in the isolation of
Neisseria meningitidis DNA sequences; and AAAB1322 to AAAB1452 represent
Neisseria meningitidis Memb polynucleotide ORF sequences, which are all
used in the exemplification of the present invention. The nucleic acid
sequences, protein sequences, and antibodies against them, can be used in
the manufacture of a composition. The composition can be used as a
medicament (or in the manufacture of a medicament) for treating,

CC preventing or diagnosing infection due to Neisserial bacteria. For
CC example, some of the identified proteins could be components of vaccines
CC against Meningococcus B; against all serotypes; and/or against all
CC pathogenic Neisseriae. Identification of sequences from the bacterium
CC will also facilitate production of biological probes, particularly
CC organism-specific probes. Attempts to make efficacious Meningococcus B
CC vaccines have failed mainly due to antigen tolerance. Multivalent
CC vaccines have also been tried but none have successfully overcome
CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
CC variable regions

CC Sequence 17782 BP; 4718 A; 4125 C; 4530 G; 4409 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,03e-147 Length: 17782
Score: 1609.50 Matches: 303
Percent Similarity: 88.19% Conservative: 48
Best Local Similarity: 76.13% Mismatches: 42
Query Match: 78.78% Indels: 5
DB: 3 Gaps: 1

US-10-039-183a-4 (1-399) x AAAB1530 (1-17782)

1 MetalatYsGluLysPheAsnArgThIrySPProHisValAsnIleGlyThrIleGlyHis 20

15097 ATGGCTAAGGAAATTTGAACGTAGCAACCGCAAGTAACGTGGACCACTGGCTCAC 15156

21 ValAspHisGlyLysThrThIleuSerAlaAlaIleSerAlaValIleuSerLeuYsgly 40

15157 GTTGACATGTGTAACCAACACTGCTGCTCTTGTGACTACTATTGTCTTAATAAATTC 15216

41 LeuAlaGluMetLysAspTYrAspAsnIleAspAsnAlaProgluIgluYarggly 60

15217 GGTGGCGCTGCAAAAGCTTATGACCAATTCGACCAACGCTCTGAAAGAAAGCTGTGT 15276

61 IleThrIleAlaThSerHisIleGluTYrGluThrGluAsnArgHisTYrAlaHisVal 80

15277 ATTACCATTTAATACCTCACACGTAGAAATACCAATGAAACCGGTCACTACGCACACGTA 15336

81 AspCysProglYHisAlaAspTYrValYIsanMetIleThrGlyAlaIaGluMetasp 100

15337 GACTGCCGGGGCAGCGCGACTACGTAAACATGATTACCGCGCGCAAAATGAGAC 15396

101 GlyAlaIleLeuValValSerAlaIaAspGlyProMetProglInThrArgIluHisIle 120

15397 GGTGCAATCTGTGATGTTCCGACGCCAGCGCCTATGCGCAAAACCGGCAACACATC 15456

121 LeuIleuSerArgGluValIglYValProHisIleValIlePheLeuAsnLysGluAspMet 140

15457 CTGCTGGCGCCGCAAGTAGCGGTACTACATCATCGTTCATATGACAAATGCGACATG 15516

141 ValAspAspGluIleuLeuGluIleuValIgluMetGluValArgGluIleuSerAla 160

15517 GTTCAGCATGCGCAGACTGTGGAATGTTGAAATGAAATTCGCGGCACTGCTGCCAGC 15576

161 TyrGluPheProglYAspAspThrProIleValAlaIglYSerAlaLeuArgAlaIleuGlu 180

15577 TACGACTTCCCGGGGAGTAGCTGCGCATGTGACAAAGTTCGCGCATAAAGCTTTGAA 15636

181 GluAlaIleValIaGlyAsnValIglYIuThrPglYgluYValIleuLysLeuMetAlaGlu 200

15637 GCGCATGCGCGT-----TACAGAGAAATTAATCTTGAACCTGCTGCCCA 15681

201 ValAspAlaTYrIleProThIleProgluArgAspThrGluYserThrPheLeuMetProVal 220

15682 TTGACACACTACATCCGACCTCCGAGGAGCGCGCAAAACGTTCTGCTGCTATC 15741

221 GluAspValPheSerIleAlaGlyArgGlyThrValValIleThrGlyArgIleGluArggly 240

15742 GAAGACGTGTTCTCATTTCCGGCGCGGTACAGTAGTAACCGGCGGTGTAGAGCGCGGT 15801

QY 241 ValValLeuValGlyValSerGluValGluIleValGlyIleArgProThrGlnLeuThrThr 260
Db 15802 ATCATCCAGCTTGTGTAACAGATGTAATCTCGCTGTAAGAAACCCAAACCACT 15861
QY 261 ValThrGlyValGluMetPheArgGlyGluLeuGluValGlyIleValSerPheVal 280
Db 15862 TGTACCGGTGTTGAATGTTCCGCAACCTGTGAGAGGTGAGCGGCGGCGCAACGTA 15921
QY 281 GlyValLeuLeuArgGlyThrIleValGlyGluValGluArgGlyMetValLeuGly 300
Db 15922 GCGATATTCCTGCGCGGTACCAACGTGAGAGGTGAGCGGTGAGATTTGGCTTAA 15981
QY 301 ProGlySerIleThrProHisIleValSerPheGluGluIleValIleThrValLeuSerIle 320
Db 15982 CCGGTACTATCTACTCTTCAACCAATTTCAAGCAAGATACCTACTGACCAAGAA 16041
QY 321 GluGlyGlyArgHisIleThrPhePheThrAsnThrArgProGlnPheThrValArgThr 340
Db 16042 GAGGGGTGTCACACTCGTTCTTCCGCACTACCGTCCGCAATTTCTACTTCCGTACC 16101
QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 16102 ACCGACGTAAACGGCGCGGTACTTGTGAGAAAGGTGAGAAATGTAAATGCCGCTGAA 16161
QY 361 AsnValIleIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrIlePhe 380
Db 16162 AACGTAAACATCCCGTAACTGATGCGCTATGCTATGAGAAAGGCTGCGCTTT 16221
QY 381 AlaIleArgGluGlyValArgThrValGlyAlaGlyValIleSerAsnIleIle 398
Db 16222 GGGATTCCGGAAAGGGCGCGTACCGTGGGCGCGGTCTTCTCTGTATC 16275

RESULT 6

Continuation (2 of 15) of AAA81490 from base 100001 (N. meningitidis B full length genom
WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aaa81490

Fragment Name	Begin	End
WP AAA81490_00	1	110000
WP AAA81490_01	100001	210000
WP AAA81490_02	200001	310000
WP AAA81490_03	300001	410000
WP AAA81490_04	400001	510000
WP AAA81490_05	500001	610000
WP AAA81490_06	600001	710000
WP AAA81490_07	700001	810000
WP AAA81490_08	800001	910000
WP AAA81490_09	900001	1010000
WP AAA81490_10	1000001	1110000
WP AAA81490_11	1100001	1210000
WP AAA81490_12	1200001	1310000
WP AAA81490_13	1300001	1410000
WP AAA81490_14	1400001	1437668

Alignment Scores:

Pred. No.:	4 356-146	Length:	110000
Score:	1609.50	Matches:	303
Percent Similarity:	88.19%	Conservative:	48
Best Local Similarity:	76.13%	Mismatches:	42
Query Match:	78.78%	Indels:	5
DB:	3	Gaps:	1

US-10-039-183a-4 (1-399) x AAA81490_01 (1-110000)

QY 1 MetAlaValGluLeuPheAsnArgThrIleProHisValAlaAsnIleGlyThrIleGlyHis 20
Db 49586 ATGGCTAAAGCAAAATTTGAGCTAGCAAAACGCACTGAACGTTGGACCAATCGCTCAC 49645
QY 21 ValAspHisGlyLeuThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuGly 40
Db 49646 GTTGCACATGTAACCACTGCTGCTTACTGCTTGTACTGTAATTTGCTAAATAATTC 49705
QY 41 LeuAlaGluMetLeuSerPheThrAspAsnIleAspAsnAlaProGluGluValGlyValGly 60

Db 49706 GGTGGCGCTGCAAAAGCTTATGACCAAAATCGACCAACGCTCTGTAAGAAAGAAAGCTCGTGT 49765
QY 61 IleThrIleAlaThrSerHisIleGlyIleValGluValGluAsnArgHisIleThrAlaIleVal 80
Db 49766 ATTACCATTAATTAATCTCACACATGTAAGATGAAATCGAAACCCGTACTACGCAACGTA 49825
QY 81 AspCysProGlyHisIleAlaAspThrValIleValAsnMetIleThrGlyValAlaIleMetAsp 100
Db 49826 GACTGCGCGGGGACGCGGCACTACGTAAACATGATTAACCGCGCGGCAAAATGAC 49885
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGlnHisIle 120
Db 49886 GGTGCAATCTGATGATGTTCCGACGCGACGCGGCTTATGCGCAACCCGCAACATC 49945
QY 121 LeuLeuSerArgGlnValGlyValProHisIleValAlaPheLeuAsnIleGlnAspMet 140
Db 49946 CTGCTGGCCCGGCAAGTAGCGGTACTTCAATCATCTGTTTCAATGAACAAATGCCACATG 50005
QY 141 ValAspAspGlnGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
Db 50006 GTTCAGATGCCGAGCTGTGAACTGTGAAATGAAATCGGACACTGCTGTCTCAGC 50065
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 50066 TACGACTTCCCGGCGATACGCGCGATTTGTAACAAGTTCCGCACTGAAAGCTTGGA 50125
QY 181 GluAlaValValGlyValAsnValGlyIleThrGlyGlyValLeuValLeuMetAlaGlu 200
Db 50126 GCGCATGCCGCT-----TACGAAGAAATAATCTTCACCTGCGTCCGCGCA 50170
QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluValThrPheLeuMetProVal 220
Db 50171 TTGACAGCTACATCCGACCTCCGACGCGCGGAGACCCGTTCTGCTGCTTATC 50230
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
Db 50231 GAAGAGCTTCTTCATTTCCGCGCGCGGTACGTACGTAAACGCGCGGTGAAGCGCGGT 50290
QY 241 ValValLeuValGlyValSerGluValGluIleValGlyIleArgProThrGlnLeuThrThr 260
Db 50291 ATCATCCAGCTTGTGTAACGATTTGAAATCGTCTGCTGTAAGAAACCAAAACCACT 50350
QY 261 ValThrGlyValGluMetPheArgGlyGluLeuGluValGlyGluValIleValAspAsnVal 280
Db 50351 TGTACCGGTGTAATGTTCCGCAACCTGTGAGAGAGTCAAGCGGCGGCGCAACGTA 50410
QY 281 GlyValLeuLeuArgGlyThrIleValGlyGluValGluArgGlyMetValLeuCysVal 300
Db 50411 GCGGTATTCCTGCGCGGTACCAACGTGAGAGAGTCAAGCGGCTCAGATTTGGCTTAA 50470
QY 301 ProGlySerIleThrProHisIleValSerPheGluGluIleThrValLeuSerIle 320
Db 50471 CCGGTACTATCTACTCTTCAACCAAAATTCAAAGCAAGATGTAACCTACTGACCAAGAA 50530
QY 321 GluGlyGlyArgHisIleThrPhePheThrAsnThrArgProGlnPheThrValArgThr 340
Db 50531 GAGGGGTGTCACACTCGTTCTTCCGCAACTACCGTCCGCAATTTCTACTTCCGTACC 50590
QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 50591 ACCGACGTAAACGGCGCGGTACTTGTGAAAGAGGTGAGAAATGTAATGCCGCGTGA 50650
QY 361 AsnValIleIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrIlePhe 380
Db 50651 AACGTAAACATACCGTGAAGATGATGCGCTTATGCGCTTATGAGAAAGAGCGCTGCTTT 50710
QY 381 AlaIleArgGluGlyValArgThrValGlyAlaGlyValIleSerAsnIleIle 398
Db 50711 GGGATTCCGGAAAGGGCGCGTACCGTGGGCGCGGTCTTCTCTGTATC 50764
RESULT 7
AAR21544

ID AAF21544 standard; DNA; 349980 BP.
XX AAF21544;
AC
XX 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.
XX Neisseria meningitidis.
XX OS
XX WO20066791-A1.
XX
XX 09-NOV-2000.
XX
XX 08-MAR-2000; 2000WO-US005928.
XX
XX 30-APR-1999; 99US-0132068P.
PR 08-OCT-1999; 99WO-US023573.
PR 28-FEB-2000; 2000GB-00004695.
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizsa M, Hickey E, Peterson J, Tetrelin H, Venter JC;
PI Mastignoli V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rapuoli R, Frazer CW, Grandi G;
XX
XX WPI, 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections.
XX
XX Claim 7, Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of Neisseria
CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
CC represent fragments of the NMB genomic sequence, as the sequence was too
CC long to go in a record on its own it was split into 8 sequences which
CC overlap each other at the beginning and end of each sequence by 49980 bp
CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
CC given in AAB5550 to AAB5853, and AAF21589 to AAF21606 represent PCR
CC primers which are used in the exemplification of the present invention.
CC The NMB genome and fragments from it have antibacterial activity, and can
CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
CC and/or antibodies which binds to the proteins can be used in compositions
CC for treating or preventing infection due to Neisserial bacteria or as a
CC diagnostic reagent for detecting the presence of Neisserial bacteria or
CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
CC computer storage medium or computer databases can be used in a search to
CC identify open reading frames (ORFs) or coding sequences within the NMB
CC genome. The DNA sequences provide further opportunities to find antigenic
CC or immunogenic proteins which are more effective in vaccines than the
CC outer membrane proteins currently used
XX
XX Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.: 1 98e-145 Length: 349980
Score: 1609.50 Matches: 303
Percent Similarity: 88.13% Conservative: 48
Best Local Similarity: 76.13% Mismatches: 42
Query Match: 78.78% Indels: 5
DB: 3 Gaps: 1

US-10-039-183a-4 (1-399) x AAF21544 (1-349980)

QY 1 MetalalysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20

DB 149586 ATGGCTAAGAAATTTGAAAGTGTCAACCGCAGTAAACGTTGGACCAATCGGTAC 149645
QY 21 ValAspHisGlyLysThrThrLeuSerAlaIleSerAlaValLeuSerLeuGly 40
DB 149646 GTTGACATGGTAAACCACTGACCTGCTGCTTGTGACTACTATTTGTCTAAATAATTC 149705
QY 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
DB 149706 GTGGCGCTGCAGAAAGCTTATGACCAATTCAGCAACGCTCTGAGAAAGAAAGCTGTGT 149765
QY 61 IleThrIleAlaThrSerHisIleGluLysGluThrGluAsnArgHisTyrAlaHisVal 80
DB 149766 ATTACCATTAATTAATCTCAACGTAAATTCGAAACCCGCTACGACACGTA 149825
QY 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGluMetAsp 100
DB 149826 GACTGCCGGGGCAGCCGACCTACTTAAACATGATTAACGGCCGACCAATGAGC 149885
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
DB 149886 GGTGCATCTCGTATGTTCCGACGCGACGGCCCTATGCGCAAAACCGGCAACATC 149945
QY 121 LeuLeuSerArgGluValGlyValProHisIleValValPheLeuAsnLysGluAspMet 140
DB 149946 CTGCTGGCCCGCAAGTGGCGTACTTACATCATCGTGTTCATGAAACAAATGCGACATG 150005
QY 141 ValAspAspGluGluLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
DB 150006 GTTCAGCATGCCGACCTGTGAACTGTGAAATGAAATTCGCGACCTGCTGCTCAGC 150065
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 150066 TACGACTTCCCGGGGAGTACGCGCGATGTGACAAAGTTCGCGACTGAAGCCTTGAA 150125
QY 181 GluAlaLysAlaGlyAsnValGlyLysIleGlyLysValLeuLysLeuMetAlaGlu 200
DB 150126 GGCGATGCCGCT-----TACGAAAGAAAACTTCGAACTGCTGCTCCGCA 150170
QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
DB 150171 TTGGACACTCATCTCCCACTCCGACGAGCCGCGGCAAAACCTTCTGCTGCTATC 150230
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
DB 150231 GAAGCGATGTTCTCATTTCCGCGCGCGGTACAGTAAACCGCGGTAGAGCGCGGT 150290
QY 241 ValValLysValGlyAspGluValGluLeuValGlyIleArgProThrGluLysThrThr 260
DB 150291 ATCATCCACGTTGGTGAAGATGAAATCGCGGTGAAAGAAACCAAAACCACT 150350
QY 261 ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal 280
DB 150351 TGTACCGGTTGTAATGTTCGCAACTCTGCGAAGGTCAGGCGGCGGCAACGTA 150410
QY 281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuGlyLys 300
DB 150411 GGCGATTTGCTGCGCGGTACCAACGTAAGTGAACGCGGTACAGTATTTGCTAAA 150470
QY 301 ProGlySerIleThrProHisIleLysPheGluGluGluLysIleThrValLeuSerLysGlu 320
DB 150471 CCGGCTACTATCACTCTCCACACCAAAATTCAGAAAGAGATATGACTGAGCAAGAA 150530
QY 321 GluGluGlyArgHisIleThrProPhePheThrAsnTyrAspProGluPheTyrValArgThr 340
DB 150531 GAGGGTGGTGGTCAACATCCGTTCTTCCCAATCCGTCGCAATTTACTCTCGTAC 150590
QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
DB 150591 ACCGACGTAAACCGCGCGGTACTTGGAGAAAGTGTGAATATGTAATGCGGTGTA 150650
QY 361 AsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrLysPhe 380

Db 150651 AACGTAACCATCAGCCTGATGTCGCTATCGCTATGAGAGAGCCTGCGCTTT 150710
 Qy 361 Aaileargylgluylgylargthvalglvalserasmetlle 398
 Db 150711 GCGATTCGGAGAGCGCGCTACCGGTGGTGGTCTTCTGTATTC 150764
 RESULT 8
 AAS51250
 ID AAS51250 standard; DNA; 1185 BP.
 AC AAS51250;
 DT 13-FEB-2002 (first entry)
 DE Enterococcus faecalis DNA for cellular proliferation protein #27.
 KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
 KW antibacterial; drug design.
 OS Enterococcus faecalis.
 PN W0200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US009180.
 XX
 PR 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;
 PI Yamamoto RT, Xu HH;
 DR WPI; 2001-611495/70.
 DR P-PSDB; AAU33391.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX
 PS Claim 27; SEQ ID NO 3832; 511bp; English.
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence encodes an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 SO Sequence 1185 BP; 377 A; 235 C; 254 G; 319 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,646-148 Length: 1185
 Score: 1608.00 Matches: 308
 Percent Similarity: 87.25% Conservative: 41

Best Local Similarity: 77.00% Mismatches: 45
 Query Match: 78.71% Indels: 6
 DB: 4 Gaps: 2
 US-10-039-183a-4 (1-399) x AAS51250 (1-1185)
 Qy 1 MetAlaLySGluYpHeaSnArGThrXyProHlValAlaSnllGlyThrlleGlyHls 20
 Db 1 ATGGCAAAAGAAAATTGACCGCTTAAATCCCATGTTAACTGATGATCGGACAC 60
 Qy 21 ValAepHlAGlyLYeThrThrlSeuSerAlaAlleSerAlValleuSerleuYegly 40
 Db 61 GTTGACCAATGTAATAACTTAACTGAGTGAATGCTGATGTAATCAAAACACGGT 120
 Qy 41 LeuAlaGluMetLYaSPryrAaPbaSnllAaPbaSnAlaProGluGluYeglyGly 60
 Db 121 GCGCGGGAAGACAAAGCTACGATTCATCGATTAACGCTCCAGAAAGAAAGAAAGCTGGA 180
 Qy 61 lIeThrlleAlaThrSerHslleGlyThrlGluThrGluAaSnArGhlleThrlAhlSval 80
 Db 181 ATCACAATCAACACTTCTCATTCGAAATGAACTGAACACTGTCACATGACACGTT 240
 Qy 81 AepCySProGlyHlsAlaAepTyVallyAaSnMetlleThrGlyAlaAglMetAap 100
 Db 241 GACTGCCAGACACGCGGACTACGTTAAACATGATGATGATGCTGCTCAATGAC 300
 Qy 101 GlyAlaIleleuValValSerAlaAepGlyProMetProGluThrArgGluHlsle 120
 Db 301 GGAGCTATCTTAAGTGTTCGTCTGCTGATGCTTATGCTCAACACGTAACATATTC 360
 Qy 121 LeuLeuSerArgGlyValAlValProHslleValAlPheLeuAaSnlySGlnAepMet 140
 Db 361 TTATTTATCAACGTAAGCTGCTTACCATATCATCGTTGATTTTAAACAAATGATAG 420
 Qy 141 ValAaPbaPegGluLeuLeuGluLeuValGluMetGlyValArgGluLeuLeuSerAla 160
 Db 421 GTTGATGACGAAGAATTATGAAATTAGTAAGAAAGAGTTGCTGATTAATATGACAA 480
 Qy 161 TyrGluPheProGlyAaPbaPheThrProHleAlaGlySerAlleuAaRglAlaGlu 180
 Db 481 TACGATTTCCAGGAGATGATGCTTCAGTTACGAGGCTTGTGTTGAACCTTTAGAA 540
 Qy 181 GluAlaLySAlaGlyAaSnValAlGlyLurPrgGlyGlyValleuYleuMetAlaGlu 200
 Db 541 GCGCAGAGCTC-----TATGAGAAAGAAATCTTGAATTAATGCTGCA 585
 Qy 201 ValAaPbaLyTrlleProThrProGluAaPbaPheThrGluYleuMetProVal 220
 Db 586 GTTGACGAATATATCCCAACTCCAGAACGATGATGACAAACCAATTCATGATGACAGTC 645
 Qy 221 GluAaPbaValPheSerlleAlaGlyArgGlyThrlValAlThrlGlyArglleGluArgGly 240
 Db 646 GAAAGCGATTCATCTCAATCACTGAGCGTACTGTTGCTACAGAGCGTGTAAACGCTGT 705
 Qy 241 ValVallySvalAlAaSPGlyValGluIleValleGlylleArg---ProThrGlnlySerThr 259
 Db 706 GAAAGTCCGCTGTGAGCAAGCTGAATCTGTTGATTAAGAGAAACATCTAAACA 765
 Qy 260 ThrValTrhGlyValAlGluMetPheArglySGluLeuGlyValGlyValAlaAaPba 279
 Db 766 ACTGTTAAGAGGTGTAAGATGTTCCGTAATTAATTAAGACTGAGAGCAGCGACAC 825
 Qy 280 ValGlyValleuLeuAaRglYThrlYblySGluGluValAlaRglYMetValleuCyS 299
 Db 826 ATCGGTGCTTAATTAAGTGTGATACAGTGAAGATGCAACGCGCAACTATTAAGCT 885
 Qy 300 LysProGlySerlleThrProHsllySlyrPheGluGlyGluIleThrValleuSerlyS 319
 Db 886 AAACCACTACATCACTCCACACAAATTCAAAGCTGAAGTATGATTAATCAAAA 945
 Qy 330 GluGluGlyGlyArgHlsThrProPhePheThrAaSnlyrArgProGluAaPbaThrValArg 339
 Db 946 GAAAGAGGCGGACGTCACATCTCATTTCACTAATCCGCTCAATTTACTTCCTGCT 1005


```

QY 241 ValValIyVAlGlyAaPgluValGluIleValGlyIleArg---ProThrGlnIySthr 259
DB 706 GAAGTTCCGGTGTGTCACCAAGTGAATCGTTGGATTAAAGACAAACATCTAAACA 765
QY 260 ThrValIThrGlyValGluMetPheArgIySgIuLeuGluIySgIuAaGlyAaPaa 279
DB 766 ACTGTTACAGGTGTTGAATGTTCCGTAATATTAATGACTAGCCTGACAGCGCACAC 825
QY 280 ValGlyValIleuLeuArgIyThrIySgIuGluIuValGluAaArgIyMetValIeuCys 299
DB 826 ATCCGGGCTTATTAATGAGTGTGACGACGGAAGATATGACGTGACAGATTAGCT 885
QY 300 LysProGlySerIleThrProHisIySlyPheGluGlyIuIleIyValIeuSerIyS 319
DB 886 AAACCAAGCTACATCTCCTCACACAAATTCAAAGCGAAGTAACTATTAATCAAAA 945
QY 320 GluGluGlyIyAArgIyThrProPhePheThrAsnIyAArgProGlnPheIyValAArg 339
DB 946 GAAGAAAGCGGAGCTACACTCCTCATTTCACTACTACCGTCCATTTCTACCTCCGT 1005
QY 340 ThrThrAspValIThrGlySerIleThrIleuProGluGlyValGluMetValMetProGly 359
DB 1006 ACAACGAGAGCTTACTGCTGTGTGAATGCGCAAGAGTACGAAATGCTAAATGCTGCT 1065
QY 360 AspAsnValIyIleThrValGluIleuIleSerProValAlaIeuGluIeuGlyThrIyS 379
DB 1066 GATTACGTTGCTATGACGTTGAATTAATTCACCCAAATGCTATTCAGAACGGAATTCCT 1125
QY 380 PheAlaIleArgGluGlyIyAArgIyThrValGlyIyAaGlyValIyValSerAsnIleIeuGlu 399
DB 1126 TTCTCTATTGTAAGGCGGAGCTACTGAGTTCAGGGCTTACTGTAATCTGTAATA 1185

RESULT 10
ADO25365
ID ADO25365 standard; DNA; 1185 BP.
XX
AC ADO25365;
XX
DT 12-AUG-2004 (first entry)
XX
DE E_faecalIs translaCtion elongaCtion faCtor Tu tuFa DNA Seq40.
XX
KW gene; d8; antimicrobIal; microbIal disease; drug compoSition; vaccine;
KW bacterIal infection; antibacterIal; food preservaCtive.
XX
OS Enterococcus faecalIs.
XX
PN WO2004041854-A2.
XX
PD 21-MAY-2004.
XX
PF 05-NOV-2003; 2003WO-CA001671.
XX
PR 05-NOV-2002; 2002US-0423757P.
PR 05-NOV-2002; 2002US-0423758P.
PR 05-NOV-2002; 2002US-0423781P.
PR 05-NOV-2002; 2002US-0423832P.
PR 05-NOV-2002; 2002US-0423875P.
PR 05-NOV-2002; 2002US-0423915P.
PR 06-NOV-2002; 2002US-0424362P.
PR 06-NOV-2002; 2002US-0424367P.
PR 06-NOV-2002; 2002US-0424370P.
PR 06-NOV-2002; 2002US-0424373P.
PR 06-NOV-2002; 2002US-0424376P.
PR 06-NOV-2002; 2002US-0424389P.
PR 07-NOV-2002; 2002US-0424502P.
PR 07-NOV-2002; 2002US-0424511P.
PR 07-NOV-2002; 2002US-0424564P.
PR 07-NOV-2002; 2002US-0424655P.
PR 08-NOV-2002; 2002US-0424686P.
PR 08-NOV-2002; 2002US-0425076P.
PR 08-NOV-2002; 2002US-0425085P.
PR 08-NOV-2002; 2002US-0425118P.

```

```

PR 08-NOV-2002; 2002US-0425126P.
PR 08-NOV-2002; 2002US-0425162P.
PR 08-NOV-2002; 2002US-0425201P.
PR 12-MAR-2003; 2003US-0453914P.
PR 12-MAR-2003; 2003US-0454021P.
PR 12-MAR-2003; 2003US-0454128P.
PR 12-MAR-2003; 2003US-0454193P.
PR 13-MAR-2003; 2003US-0454215P.
PR 13-MAR-2003; 2003US-0454218P.
PR 13-MAR-2003; 2003US-0454487P.
PR 13-MAR-2003; 2003US-0454507P.
PR 13-MAR-2003; 2003US-0454536P.
PR 14-MAR-2003; 2003US-0455010P.
PR 14-MAR-2003; 2003US-0455036P.
PR 14-MAR-2003; 2003US-0455054P.
PR 14-MAR-2003; 2003US-0455082P.
PR 17-MAR-2003; 2003US-0455191P.
PR 17-MAR-2003; 2003US-0455192P.
PR 17-MAR-2003; 2003US-0455334P.
PR 17-MAR-2003; 2003US-0455335P.
PR 17-MAR-2003; 2003US-0455343P.
PR (AFPI-) AFPIINUM PHARM INC.
XX
PA Edwards A, Dharamsi A, Vedadi M, Domagala M, Nethery K;
PI Mansoury K, Pinder B, Alam MZ, Ng I, Vireg C, Houston S;
PI McDonald M, Buzadzija K;
XX
DR WPI; 2004-400642/37.
XX
DB P-PSDB; ADO25366.
XX
PT Bacterial polypeptide composition useful for treating bacterial
PT infection, has isolated, recombinant bacterial polypeptide such as GTP-
PT binding protein Era from Pseudomonas aeruginosa or adenylsuccinate lyase
PT from Enterococcus faecalis.
XX
PS Claim 28; SEQ ID NO 40; 566bp; English.
XX
XX
CC This invention relates to the identification of novel protein targets for
CC the development of antimicrobial drugs against pathogenic bacteria.
CC Specifically, it refers to recombinant proteins derived from
CC Staphylococcus aureus, Helicobacter pylori, Streptococcus pneumoniae,
CC Escherichia coli, Enterococcus faecalis or Pseudomonas aeruginosa. The
CC present invention describes providing a three-dimensional structure for
CC these crystallised proteins to identify a potential modulator for the
CC prevention or treatment of microbial diseases. Furthermore, connecting a
CC protein with a modulator can be useful for assaying protein activity and
CC hence its viability in drug composition or vaccine. Accordingly, such
CC compositions can be useful for treating bacterial infections, developing
CC antibacterial agents useful as food preservatives or treating food
CC products to eliminate potential pathogens. This polynucleotide sequence
CC is a bacterial DNA encoding a protein target of the invention.
XX
SQ Sequence 1185 BP; 376 A; 235 C; 254 G; 320 T; 0 U; 0 Other;
XX
XX
Alignment Scores:
Pred. No.: 1,64e-148 Length: 1185
Score: 1608.00 Matches: 308
Percent Similarity: 87.25% Conservative: 41
Best Local Similarity: 77.00% Mismatches: 45
Query Match: 78.71% Indels: 6
DB: 12 Gaps: 2
US-10-039-183a-4 (1-399) x ADO25365 (1-1185)
QY 1 MetAlaIyGluIyPheAsnArgThrIySProHisValAsnIleGlyThrIleGlyHis 20
DB 1 ATGCGAAAAGAAAATTTCACCGTCTTAATCCCATGTTAAACATTCGATTCGACAC 60
QY 21 ValAaPheIleGlyIySthrIleuSerAlaIleIleSerAlValIeuSerIeuIySgIy 40
DB 61 GTTGACCAATGTAAGAACTAATTAACAGCTGCAATGCTACTGATTAATCAAAACACGGT 120

```

```

QY 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
DB 121 GCGCGGAGACAAAGCTACGATTCATCGATTAACGCTCCAGAGAAAAGAACGTGGA 180
QY 61 ILeThrIleAlaTherSerIleIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
DB 181 ATCAACAAATCAACATTCATCGAATATGAAACGAAACCTGCACATACACACGCT 240
QY 81 AAspCysProGluYHisAlaAspTyrValLysAsnMetIleThrGluValAlaIleGluMetAsp 100
DB 241 GACTGCCCGAGACATCCGACATACGTTAAACATATACATCGTGCCTCTCAATATGAC 300
QY 101 GluAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
DB 301 GGAGCTATCTTAGTAGTTCCTGCTGCTGATGGCTATGCTCAACACATGTAACATATC 360
QY 121 LeuLeuSerArgGluValGluValProHisIleValAlaPheLeuAsnLysGluAspMet 140
DB 361 TTAATATACAGTAACGTTGCTGTATACCATCATCGTGTATTCCTTAAACAAATGATATAG 420
QY 141 ValAspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
DB 421 GTTGATGACGAAGAAATTTAGAAATTAAGTAAGAAATGAAAGTTGCTGACTTATTATCAGA 480
QY 161 TyrGluPheProGluYAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 481 TACGATTTCCAGGCGCATGATGTTCCATTCGCGAGGTTCTGCTTGAAGCTTTAGAA 540
QY 181 GluAlaLysAlaGluYAsnValGluLysTyrGluLysValLeuLeuMetAlaGlu 200
DB 541 GCGCAGCACTCT-----TATGAAGAAAATACTTGAATTAATGCTGCA 585
QY 201 ValAspAlaTyrIleProThrProGluYAspAspThrGluLysThrPheLeuMetProVal 220
DB 586 GTTGACGAATATATCCCACTCCAGACCTGATACGAAACCAATTCATGATGCCAGTC 645
QY 221 GluAspValPheSerIleAlaGluArgGlyThrValValThrGluYArgIleGluArgGly 240
DB 646 GAAGACGATTCCTCAATCACTGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705
QY 241 ValValLysValGluYAspGluValGluIleValGluIleArg---ProThrGluLysThr 259
DB 706 GAAGTTCCGGTGGTGACGAAAGTTGAAATCGTTGATTAAGAACGAAACATCTAAACAA 765
QY 260 ThrValIleThrGluValGluMetPheArgLysGluLeuGluGlyGluLysAlaGluAspAsn 279
DB 766 ACTGTTACAGGTGTTGAATGTTCCGTAATATTATTAAGCTACGCTGACAGGCGACACAC 825
QY 280 ValGluValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCys 299
DB 826 ATCCGGGCTTTATTAAGTGTGTAGACGTAAGATATCGAAGCTGCAACAGTATTAAGCT 885
QY 300 LysProGlySerIleThrProHisLysLysPheGluGluIleLysTyrValLeuSerLys 319
DB 886 AAACACGACATCACTCCACACACAAATTCAAAGCGAAGTAAATCAATATATCAAAA 945
QY 320 GluGluGluGlyYArgHisThrProPhePheThrAsnTyrArgProGluPheTyrValArg 339
DB 946 GAAGAAAGGCGACATCACTCATCTTCACTACTACCGCTCAATTCATCTCCGT 1005
QY 340 ThrThrAspValThrGlySerIleThrLeuProGluGluValGluMetValMetProGly 359
DB 1006 ACAACAGACGTTACTGCTGTGTGAGAAATGCCAAGAGTACGAATATGCTGCTGCT 1065
QY 360 AspAsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLys 379
DB 1066 GATTAACGTTGATGACGCTTAATTAATCAACCAATGCGATACAAACGAAACTCGT 1125
QY 380 PheAlaIleArgGluGluYArgThrValGluValAlaGluValValSerAsnIleIleGlu 399
DB 1126 TTCTCTATTCTGTAAGGCGACGTAAGCTGTAAGGCTGTTGTTACTGAATTCGTTAAA 1185

```

```

AAS52748
ID AAS52748 strand; DNA; 1188 BP.
XX
AC AAS52748;
XX
DT 13-FEB-2002 (first entry)
XX
DE Enterococcus faecalis DNA for cellular proliferation protein #176.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Enterococcus faecalis.
XX
PN W0200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT- ) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GI;
PI Yamamoto RT, Xu HH;
XX
DR MPI: 2001-611495/70.
XX
DR P-PSDB; AAU34889.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 6385; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins. The proteins can
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1188 BP; 379 A; 235 C; 254 G; 320 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1,656-148 Length: 1188
Score: 1608.00 Matches: 308
Percent Similarity: 87.25% Conservative: 41
Best Local Similarity: 77.00% Mismatches: 45
Query Match: 78.71% Indels: 6
DB: 4 Gaps: 2
QY US-10-039-183a-4 (1-399) x AAS52748 (1-1188)
1 MetAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20

```

```

Db      1 ATGCAAAAGAAAATTGACCGTTCTTAATCCCATGTTACATGTTACTATCGGACAC 60
Qy      21 ValaaphhlglylvthrthrleuSerAlaAlaileSerAlaValleuSerleuylsgly 40
Db      61 GTTGACCAATGTGMAAACTACATTAACGCTGCAATGCTACTGTTATTAACAAACACGGT 120
Qy      41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluylsgly 60
Db      121 GCGGGGGAAGCAACAAAGCTACGATTCATCGATTACGCTCCAGAGAAAAGAAAGCTGGA 180
Qy      61 IlerhrllealThrSerhlsileglutryglutrhgluAsnArgHlsTyrAlaHlsVal 80
Db      181 ATCAATCAATCAACACTTCTCATATCGAATATGAAGAACTGAAACCTGCATCATCAACGGT 240
Qy      81 AspCyPProGlyHlsAlaAspTyrValylsAsnMetIleThrGlyAlaAlaGluMetAsp 100
Db      241 GACTGCCCAAGGACAGCGGACTACGTTAAACATGATCATCGTGTGCTGCTCAATGAC 300
Qy      101 GlyAlaIleLeuValValSerAlaAlaAspGlyPrometProGluThrArgLuhlsIle 120
Db      301 GAGCTATCTTATGTTAGTTCTGCTGATGCTCTATGCTCTCAACGCTCAACATATC 360
Qy      121 LeuLeuSerArgGlnValGlyValProhlsileValIlePheLeuAsnLysGlnAspMet 140
Db      361 TTATTATACACGTAAGCTGTGTGATCATCATCGTTGTTATTTAAACAAATGATATG 420
Qy      141 ValaAspArgGlnGluLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
Db      421 GTTGATGACGAAAGATTATTAAGATTAGAGAAAGAAAGTGGACGCTTATTAACAGAA 480
Qy      161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db      481 TACGATTCGCCAGGCGATGATCTTCATGATTCGAGGCTGCTTGAACACCTTAGAA 540
Qy      181 GluAlaLysAlaGlyAsnValGlyLysGluLysValLeuLysLeuMetAlaGlu 200
Db      541 GCGCAGCAGACTC-----TATGAAGAAAATCTTGAATATATGCTGCA 585
Qy      201 ValaAspAlaTyrIleProThrProGluLysArgPheThrPheLeuMetProVal 220
Db      586 GTTGACGATATATATCCCAATCCAGAACCTGATGACGAAACCAATTCATGATGCGACATC 645
Qy      221 GluAspValPheSerIleAlaGlyArgGlyThrValIleThrGlyArgLysGluArgGly 240
Db      646 GAAAGCGATTTCTCATCTGAGCGTGTGCTGTTGCTACAGGACGTCGTAAACGTGT 705
Qy      241 ValValLysValGlyAspGluValGluIleValGlyIleArg---ProThrGlnLysThr 259
Db      706 GAAGTTCCGCTGTGACGAAAGTTGAAATCGTTGTTTAAAGACGAAACATCTAAACA 765
Qy      260 ThrValIleThrGlyValGluMetPheArgLysGluLeuGlyValGluIleValAspAsn 279
Db      766 ACTGTTACAGGCTGTGAATGTTCCGTAATATTAATGACTGCGAAGCAGGACGACAAAC 825
Qy      280 ValGlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuGly 299
Db      826 ATCGGTCTTATTAATGCTGTGTACACGTAAGATATGAAACGTAAGAAATATTAAGT 885
Qy      300 LysPProGlySerIleThrProhlsLysLysPheGluGluIleTyrValLeuSerLys 319
Db      886 AAACAGCATCAATCACTCCACACACAAATTCAAAGCAAGTATACATATTAACAAA 945
Qy      320 GluGluGlyGlyValArgHlsThrProPhePheThrAsnTyrArgProGluPheTyrValArg 339
Db      946 GAAGAGAGCGGACGTCACACTCCATTTCTCACTAACCGTCCCAATTTACTACTCGT 1005
Qy      340 ThrThrAspValThrGlySerIleThrLeuPProGluGlyValGluMetValMetProGly 359
Db      1006 ACAACAGACGTTACTGCTTTTAAATTTGCAAGAGGATGAAATGTAATGCTGTGT 1065
Qy      360 AspAsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrLys 379

```

```

Db      1066 GATTAACGTTGCTATGACGCTTGAATTAATTCACCAATCGCTATGAAAGACGAACTCGT 1125
Qy      360 PheAlaIleArgGluGlyValArgThrValGlyAlaGlyValIleSerAlaIleIleGlu 399
Db      1126 TTCTCTATTCGTAAGGCGGACGTAAGTATGAGGCTGAGGCTGTGTAATCGTTAAA 1185

RESULT 12
ID      ACA41574 standard, DNA, 1185 BP.
XX
AC      ACA41574;
XX
DT      19-JUN-2003 (first entry)
XX
DE      Prokaryotic essential gene #23231.
XX
KW      Antisense; ds; prokaryotic essential gene; cell proliferation;
KW      drug design; gene.
XX
OS      Neisseria meningitidis.
XX
PN      WO200277183-A2.
XX
PD      03-OCT-2002.
XX
PF      21-MAR-2002; 2002MO-US009107.
XX
PR      21-MAR-2001; 2001US-00815242.
PR      06-SEP-2001; 2001US-00948993.
PR      25-OCT-2001; 2001US-0342923P.
PR      08-FEB-2002; 2002US-00072851.
PR      06-MAR-2002; 2002US-0362699P.
XX
PA      (ELITR-) ELITRA PHARM INC.
XX
PI      Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI      Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR      MPI: 2003-029926/02.
XX
DR      P-PSDB; ABU37704.
XX
PT      New antisense nucleic acids, useful for identifying proteins or screening
PT      for homologous nucleic acids required for cellular proliferation to
PT      isolate candidate molecules for rational drug discovery programs.
XX
PS      Claim 14; SEQ ID NO 29444; 1766bp; English.
XX
CC      The invention relates to an isolated nucleic acid comprising any one of
CC      the 623 antisense sequences given in the specification where expression
CC      of the nucleic acid inhibits proliferation of a cell. Also included are:
CC      (1) a vector comprising a promoter operably linked to the nucleic acid
CC      encoding a polypeptide whose expression is inhibited by the antisense
CC      nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC      polypeptide or its fragment whose expression is inhibited by the
CC      antisense nucleic acid; (4) an antibody capable of specifically binding
CC      the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC      proliferation or the activity of a gene in an operon required for
CC      proliferation; (7) identifying a compound that influences the activity of
CC      the gene product or that has an activity against a biological pathway
CC      required for proliferation, or that inhibits cellular proliferation; (8)
CC      identifying a gene required for cellular proliferation or the biological
CC      pathway in which a proliferation-required gene or its gene product lies
CC      or a gene on which the test compound that inhibits proliferation of an
CC      organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC      compound's activity; (11) a culture comprising strains in which the gene
CC      product is overexpressed or underexpressed; (12) determining the extent
CC      to which each of the strains is present in a culture or collection of
CC      strains; or (13) identifying the target of a compound that inhibits the
CC      proliferation of an organism. The antisense nucleic acids are useful for
CC      identifying proteins or screening for homologous nucleic acids required
CC      for cellular proliferation to isolate candidate molecules for rational
CC      drug discovery programs, or for screening homologous nucleic acids
CC      required for proliferation in cells other than S. aureus, S. typhimurium,

```


CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 1185 BP; 315 A; 321 C; 302 G; 247 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,31e-148	Length:	1185
Score:	1606.50	Matches:	302
Percent Similarity:	88.19%	Conservative:	49
Best Local Similarity:	75.88%	Mismatches:	42
Query Match:	78.63%	Indels:	5
	8	Gaps:	1

US-10-039-183a-4 (1-399) x ACA41574 (1-1185)

```
QY 1 MetAlaLySGluLyPhehAnArGThrLySPRoHISvaLAsnIlleGlyThrIlleGlyHis 20
Db 1 ATGGCTTAAGAAAATTGCAAGCTAGCAAAACCGCAAGTAAAGCTTGCGACCATCGGTAC 60
QY 21 ValAspHisGlyLyThrThreUsSerAlaAlleSerAlValleUsSerLeuLySgly 40
Db 61 GTTGAACATGTGTAAACCAACCTGACCTGCGCTTGACTACTATTGCTAAATAATTC 120
QY 41 LeuAlaGluMetLyAspLyAspLyAspSerAlaAspGluGluGluLySgluLyArgLy 60
Db 121 GCGCGTGTCTGCAAAAGCTTACGAAATCGAACAGCACCGCAAGAAAACACAGCGGT 180
QY 61 IleThrIleAlaThSerHisIleGlyLyThrGluThrGluAsnArGHisIleLyAlaHisVal 80
Db 181 ATTACCATTAACACTCGACAGCTGGAATCGAAACCGAACCCGCACTACGACACAGTA 240
QY 81 AspCySPProGlyHisIleAlaAspLyValLyAsnMetIleThrGlyAlaAlaGluMetAsp 100
Db 241 GACTGTCCGGGGGACGCCCACTACCTTAAACATGATTAACCGGGCGACACAAATGAGC 300
QY 101 GlyAlaIleLeuValValSerAlaAlaAspLyPromeProGluThrArgLyHisIle 120
Db 301 GGTGCAATCTGTGATGTTCCGAGCCGAGCGGCTTATGCCCCCAACCCGCAACACATC 360
QY 121 LeuLeuSerArgGluValGlyValProHisIleValAlaPheLeuAsnLySgluAspMet 140
Db 361 CTGCGTGGCCCGTCAGTAGTAGGCGGTACCTTATCATCATGCTTATGAAACAAAGCCAGC 420
QY 141 ValAspAspGluGluLeuLeuGluLeuValGluMetGluValAlaArgLyLeuLeuSerAla 160
Db 421 GTCCGACGATGCCGAGACTGTGTGAACCTGTGAAATCGGAGACTGCTGTCCAGC 480
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 481 TACGACTTCCCGGAGACGACTGCCGATCGTACAAAGTTCCGACACTGAAACCTTGGA 540
QY 181 GluAlaLySAlaGlyAsnValGlyGluThrGlyGlyLySValLeuLyLeuMetAlaGlu 200
Db 541 GCGCATGCCGCT-----TACGAAGAAAATAATTGGAATGGCTGCTGCA 585
QY 201 ValAspAlaLyThrIleProThrProGluArgAspThrGluLyThrPheLeuMetProVal 220
Db 586 TTGGACAGCTACATCCCGACTCCCGAGCGTCCGTCGCAAAACCTTTCTTGTGCTTATC 645
QY 221 GluAspValPheSerIleAlaGlyArgLyThrValAlaThrGlyArgLyLeuLyArgLy 240
Db 646 GAAAGCGATTTCTATTTCGCGTGTGTACAGTAGTAAACCGTGTGTACAGCGCGGT 705
QY 241 ValValLySValGlyAspGluValGluIleValGlyIleArgProThrGluLySThrThr 260
Db 706 ATCATCCACAGCTGTGAGACGAGATCGAATCGTCCGTCTGAAAGAAACTCAAAAACCACT 765
QY 261 ValThrGlyValGluMetPheArgLySgluLeuGluLySgluValAlaGlyAspAsnVal 280
Db 766 TGTACCGGTGTGGAATGTTCCGCAAAACTGTGACGAAGGTCGAAGCGGCAACAGTA 825
```

```
QY 281 GlyValLeuLeuArgGlyThrLySgluGluValGluArgGlyMetValLeuCySlys 300
Db 826 GCGGATTCCTCGCGGATACCAACGTGAAGACGTAGAGCGTGTGCAAGTATGGCTAAA 885
QY 301 ProGlySerIleThrProHisIleLySAspPheGluGlyIleLyThrValleUsSerLySglu 320
Db 886 CCGGGTACAAATCATCTCTCACACCAAGTTCAAGACGAAGTATAGTACTGAGCAAGAA 945
QY 321 GluGlyLyArgHisIleThrPhePheThrAsnLyPAspProGluPheLyThrValArgThr 340
Db 946 GAGGGCGGCGGCATACCCCGTCTTCCGCAACTACCGTCCCAATTTACTTCCGTAC 1005
QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 1006 ACCGACGTACCGGGCGGCTTACTTTGGAAGAAGGTGTGAAATGTGTATATGCCGCGAG 1065
QY 361 AsnValLySleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLySphe 380
Db 1066 AACGTAAACCATCACCGTGAACCTGATTCGCTTACCTATGCAAGAAAGTTGCGCTTT 1125
QY 381 AlaIleArgGlyGlyLyArgThrValGlyValAlaSerAsnIleIle 398
Db 1126 GCGATTCCGGAAGCGCGCTTACCGTGTGCGCGGTCTTCTGTATTC 1179
RESULT 13
ACA41586
ID ACA41586 standard; DNA; 1185 BP.
XX
XX ACA41586;
AC
XX
XX 19-JUN-2003 (first entry)
DT
XX
XX DE Prokaryotic essential gene #23243.
XX
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX OS Neisseria meningitidis.
XX
XX PN W0200277183-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 21-MAR-2002; 2002MO-US009107.
XX
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR MPI; 2003-029926/02.
XX DR P-PSDB; AB037716.
XX
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX PS Claim 14; SEQ ID NO 29456; 1766pp; English.
XX
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX the 623 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
```


PI Rappuoli R, Pizza M;
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used
PT in the diagnosis and treatment of *N. meningitidis* infection and other
PT *Neisseria* infections, for example, *N. gonorrhoea*.

Claim 7; Page 1331-1350; 1760pp; English.

CC The present invention describes methods of obtaining immunogenic proteins
CC from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent
CC specifically claimed *Neisseria meningitidis* genomic DNA sequences;
CC AAA81260 to AAA81303 and AAA825620 to AAA825663 represent *Neisseria* DNA
CC sequences and their corresponding proteins; AAA81254 to AAA81259 and
CC AAA81304 to AAA81321 represent PCR primers used in the isolation of
CC *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent
CC *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all
CC used in the exemplification of the present invention. The nucleic acid
CC sequences, protein sequences, and antibodies against them, can be used in
CC the manufacture of a composition. The composition can be used as a
CC medicament (or in the manufacture of a medicament) for treating,
CC preventing or diagnosing infection due to *Neisseria* bacteria. For
CC example, some of the identified proteins could be components of vaccines
CC against *Meningococcus B*; against all serotypes; and/or against all
CC pathogenic *Neisseriae*. Identification of sequences from the bacterium
CC will also facilitate production of biological probes, particularly
CC organism-specific probes. Attempts to make efficacious *Meningococcus B*
CC vaccines have failed mainly due to antigen tolerance. Multivalent
CC vaccines have also been tried but none have successfully overcome
CC antigenic variability. The provision of further, complete sequences may
CC provide an opportunity to identify secreted or surface exposed proteins
CC that may be presumed targets for the immune system and which are not
CC antigenically variable or at least more conserved than other more
CC variable regions

XX SQ Sequence 65632 BP; 16704 A; 16327 C; 15474 G; 17126 T; 0 U; 1 Other;

XX Alignment Scores:

Pred. No.: 4,396-146 Length: 65632
Score: 1606.50 Matches: 302
Percent Similarity: 88.19% Conservative: 49
Best Local Similarity: 75.88% Mismatches: 42
Query Match: 78.63% Indels: 5
DB: 3 Gaps: 1

US-10-039-183A-4 (1-399) x AAA81502 (1-65632)

QY 1 MetalAlYVGluLYsPhaAaRgThRySPRoHISValASnIleGlyThRtleGlyHis 20
DB 1901 ATGGCTAAGGAAATAATTCGACGTAAGCAACCGACGTAACGTTGGACCATCGGTCA 1842
QY 21 ValAaPHISGlyLYsThRleuSerIleAaIleSerIleValleuSerleuYsgly 40
DB 1841 GTTGACCAATGTGAAACCAACCTGACCTGCTTGTGACTACTATTGCTTAAATAATTC 1782
QY 41 LeuAlaGluMetLYsAspTYsAspAsnIleAspAsnAlaProGluGluYsglyuArgly 60
DB 1781 GCGGCTGCTGCAAAAGCTTACGACCAATCGACCAACGACCCGAGAAAGAAACGACGCGGT 1722
QY 61 IleThRtleAaThSerHisIleGlyuArglyuThRguAsnAArgHisTYsRlaHisVal 80
DB 1721 ATTACCATTTAACACCTCGCAGCGTGAATTCGAAACCGAAACCGGACATGACACACGTA 1662
QY 81 AspCYsPProGlyHisAlaAspTYsValLYsAaMetIleThGlyValAlaGluMetAsp 100
DB 1661 GACTCCCGGGGACAGCGACGACGTAACGTTAAACATGATTAACGCGGCGCCACCAATATGAC 1602
QY 101 GlyAlaIleLeuValValSerIleAaAspGlyProMetProGluThRArglyuHisIle 120
DB 1601 GGTGCAATCTGTGTATGTTCCGACGACGAGCGCTTATGCGCAAAACCGCAACACATC 1542
QY 121 LeuLeuSerArgGluValGlyValProHisIleValIlePheLeuAsnLYsGluAspMet 140

DB 1541 CTGCTGCGCCCGCAAGTAGCGCTTACCTTACATCATCATCGTTTCATACCAAAAGCACACAG 1482
QY 141 ValAaPAspGluGluLeuGluValGluMetGluValArgGluLeuLeuSerIle 160
DB 1481 GTTGCAGATGCCGAGCTGTGGAACTGGTGAATTCGAAATTCGGACCTGCTGCCAC 1422
QY 161 TyrGluPheProGlyAspAspThRProIleValAlaGlySerIleAaValArgAlaLeuGlu 180
DB 1421 TACGACTTCCCGGGAGATGATCGCCGATTTGTAACAAGGTTCCGACCTGAAACCTTGGA 1362
QY 181 GluAlaLYsAlaGlyAsnValGlyuTrpGlyGlyLYsValleuLYsLeuMetAlaGlu 200
DB 1361 GCGCATGCCGCT-----TACGAAGAAAAAATTTGCAACTGCTCCGCA 1317
QY 201 ValAaPAlaTYsIleProThRProGluArgAspThRGuLYsThRPhelMetProVal 220
DB 1316 TTGGACAGCTACATCCCACTCCGAGGAGCGCGTGACAAACCGTTCTGCTGCTATC 1257
QY 221 GluAaPValPheSerIleAlaGlyArglyThRValValThRGuLYsArgIleGluArgly 240
DB 1256 GAAGACGTTCTTCATTTCGCGCGCGGTACAGTAAACCGGCGGTAGAGCGCGGT 1197
QY 241 ValValLYsValGlyAspGluValGluIleValGlyIleArgProThRGuLYsThRThr 260
DB 1196 ATCATCCACGTTGTGACGAGATTGAATTCGTGCTGGAAGAAACCAAAACCACT 1137
QY 261 ValThRGuLYsAlaGluMetPheAaRgLYsLeuGluGlyuValGlyAlaGlyAspAsnVal 280
DB 1136 TGTACCGGTTGTAATGTTCCGCAAACTGTCGCAAGGTCAGCGCGGACACACGTA 1077
QY 281 GlyValleuLeuArgGlyThRLYsLYsGluGluValGluArgLYsMetValleuCYsLYs 300
DB 1076 GCGGATTTGCTGCGGAGTACCAACGTAAGAAGCGTAACGCGTATGCTTAA 1017
QY 301 ProGlySerIleThRProHisLYsLYsPheGluGlyuIleTYsValleuSerLYsGlu 320
DB 1016 CCGGCTACTATCATCTCCACACCAAAATTCAGAAAGAGTAATGACTGACGCAAGAA 957
QY 321 GluGlyLYsArgGlyThRProPheThRAsnTYsArgProGluPheTYsValArgThR 340
DB 956 GAGGATGTCGTCACACTCCGTTCTTCCCACTACCGTCCCAATTTACTTCCGTACC 897
QY 341 ThrAaPValThRGuLYsSerIleThRleuPProGluGlyValGluMetValMetProGlyAsp 360
DB 896 ACCGACGTAAACCGGCGGCTTACTTGAAGAAGGTAGAAATGCTATATGCCGCTAA 837
QY 361 AsnValLYsIleThRValGluLeuIleSerProValAlaLeuGluLeuGlyThRysPhe 380
DB 836 AACGTAAACCATCACCGTGAACCTGATTCGCTATGCGTATGGAAGAAAGGCTCGCTTT 777
QY 381 AlaIleArgGluGlyLYsArgThRValGlyValAlaGlyValIleSerAsnIleIle 398
DB 776 GGAATTCGGAAGGCGCGGTACCGTGGTCCGCGGTTCCTTCTGTATC 723
RESULT 15
ADO25367
ID ADO25367 standard; DNA; 1185 BP.
XX ADO25367;
XX 12-AUG-2004 (first entry)
XX E_faecalis translation elongation factor Tu tufa DNA Seq42.
XX gene; ds; antimicrobial; microbial disease; drug composition; vaccine;
XX bacterial infection; antibacterial; food preservative.
XX Enterococcus faecalis.
XX OS
XX PN WO2004041854-A2.
XX 21-MAY-2004.

XX	05-NOV-2003; 2003WO-CA001671.
PF	
XX	
PR	05-NOV-2002; 2002US-0423757P.
PR	05-NOV-2002; 2002US-0423758P.
PR	05-NOV-2002; 2002US-0423791P.
PR	05-NOV-2002; 2002US-0423832P.
PR	05-NOV-2002; 2002US-0423875P.
PR	05-NOV-2002; 2002US-0423915P.
PR	06-NOV-2002; 2002US-0424367P.
PR	06-NOV-2002; 2002US-0424370P.
PR	06-NOV-2002; 2002US-0424373P.
PR	06-NOV-2002; 2002US-0424376P.
PR	06-NOV-2002; 2002US-0424389P.
PR	07-NOV-2002; 2002US-0424502P.
PR	07-NOV-2002; 2002US-0424651P.
PR	07-NOV-2002; 2002US-0424664P.
PR	07-NOV-2002; 2002US-0424665P.
PR	08-NOV-2002; 2002US-0424968P.
PR	08-NOV-2002; 2002US-0425076P.
PR	08-NOV-2002; 2002US-0425085P.
PR	08-NOV-2002; 2002US-0425118P.
PR	08-NOV-2002; 2002US-0425126P.
PR	08-NOV-2002; 2002US-0425162P.
PR	08-NOV-2002; 2002US-0425201P.
PR	12-MAR-2003; 2003US-0453914P.
PR	12-MAR-2003; 2003US-0454021P.
PR	12-MAR-2003; 2003US-0454128P.
PR	12-MAR-2003; 2003US-0454193P.
PR	13-MAR-2003; 2003US-0454215P.
PR	13-MAR-2003; 2003US-0454218P.
PR	13-MAR-2003; 2003US-0454487P.
PR	13-MAR-2003; 2003US-0454507P.
PR	13-MAR-2003; 2003US-0454536P.
PR	14-MAR-2003; 2003US-0455010P.
PR	14-MAR-2003; 2003US-0455036P.
PR	14-MAR-2003; 2003US-0455054P.
PR	14-MAR-2003; 2003US-0455082P.
PR	17-MAR-2003; 2003US-0455191P.
PR	17-MAR-2003; 2003US-0455192P.
PR	17-MAR-2003; 2003US-0455344P.
PR	17-MAR-2003; 2003US-0455355P.
PR	17-MAR-2003; 2003US-0455343P.
XX	
PA	(AFPI-) AFFININUM PHARM INC.
XX	
FI	Edwards A, Dharamsi A, Vedadi M, Domagala M, Nethery K,
PI	Mansoury K, Pinder B, Alam MZ, Ng I, Vireg C, Houston S;
PI	McDonald M, Buzadzija K;
XX	
DR	WPI 2004-400642/37.
XX	
XX	P-P8DB; ADO25368.
PT	
PT	Bacterial polypeptide composition useful for treating bacterial
PT	infection, has isolated, recombinant bacterial polypeptide such as GMP-
PT	binding protein Era from Pseudomonas aeruginosa or adenylsuccinate lyase
PT	from Enterococcus faecalis.
XX	
XX	Claim 28; SEQ ID NO 42; 566pp; English.
PS	
XX	
CC	This invention relates to the identification of novel protein targets for
CC	the development of antimicrobial drugs against pathogenic bacteria.
CC	Specifically, it refers to recombinant proteins derived from
CC	<i>Staphylococcus aureus</i> , <i>Helicobacter pylori</i> , <i>Streptococcus pneumoniae</i> ,
CC	<i>Escherichia coli</i> , <i>Enterococcus faecalis</i> or <i>Pseudomonas aeruginosa</i> . The
CC	present invention describes providing a three-dimensional structure for
CC	these crystallised proteins to identify a potential modulator for the
CC	prevention or treatment of microbial diseases. Furthermore, contacting a
CC	protein with a modulator can be useful for assaying protein activity and
CC	hence its viability in drug composition or vaccine. Accordingly, such
CC	compositions can be useful for treating bacterial infections, developing
CC	antibacterial agents useful as food preservatives or treating food

CC products to eliminate potential pathogens. This polynucleotide sequence
CC is a bacterial DNA encoding a protein target of the invention.
xx
50 Sequence 1185 BP; 376 A; 234 C; 254 G; 321 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,01e-147	Length:	118
Score:	1600.00	Matches:	307
Percent Similarity:	86.75%	Conservative:	40
Best Local Similarity:	76.75%	Mismatches:	47
Query Match:	78.32%	Indels:	6
DB:	12	Gaps:	2

US-10-039-183A-4 (1-399) x ADO25367 (1-1185)

1 MeAlLlyeGluLybPheAsnArgThLyPProHlsvAlaInlIeGlyThIleGlyVhlP
Db 1 ATGGCAAAAGAAAAATTGACCGGTTCTAAATCCCATGTTAAACATTGGTACATTCGACAC 60
Oy 21 ValAspHisGlyLyLeuThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLyGly 40
Db 61 GTTGACCAATGCTAAATCTACATTAAACAGTCGCAATTCCTACTGATTATCAAAACAGCGT 120
Oy 41 LeuAlaGluMetLyAspArgThrAspAsnIleAspAsnAlaProGluGluLybGluArgLy 60
Db 121 GCGCGGAAAGCACAAAAGCTACATTCCTATCGATAAACGCTCCAGAAAGAAAAGAACGTGGA 180
Oy 61 IleThrIleAlaIleSerHisIleGluLyArgIuThrGluAsnArgHisArgAlaHisVal 80
Db 181 ATCAACATCAACACCTTATCATATCGAATATGAAATGAAATCGGTCACTATGACACAGTT 240
Oy 81 AspCysArgProGlyVhlValAspTyrValLybAsnMetIleThrGlyAlaIleGluMetAsp 100
Db 241 GACTGGCCCAAGACATCGACATCGATCACTTAAAAACAAGAACACTGGGTGCTCAAAATGAC 300
Oy 101 GluValIleLeuValValSerAlaAlaAspGlyProMetProGluIleThrArgGluHisIle 120
Db 301 GAGACTATCTTAGTAGTTCTCTGCTGATGGTCTTATGCTTCAACACATGAAACATATTC 360
Oy 121 LeuLeuSerArgGluValGlyValProHisIleValPheLeuAsnLyGluAspMet 140
Db 361 TTATTATACAGTAAACGTTGGTATCAATACATCGTTGATTCTTAAACAAATAGGATATG 420
Oy 141 ValAspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
Db 421 GTTGAATGACGAAGAAATATTGAATTAGACGAAGAAAGAAATCGTACATTATATTCGAA 480
Oy 161 TyrGluPheProGlyAspAspThrProIleValIleArgIleSerAlaLeuArgAlaLeuGlu 180
Db 481 TAGCATTTCCACAGCGCATGATGTTCCACAGTATCCGACAGTTCTGCTTGAAGATTGAA 540
Oy 181 GluAlaLybAlaGlyAsnValGlyGluTProGlyGluLybValLeuLybLeuMetAlaGlu 200
Db 541 GCGCAAGAGACT-----TATGAAGAAAAAATCTTGAATATATGCTGCA 585
Oy 201 ValAspAlaTyrIleProThrProGluLybArgAspThrGluLybThrPheLeuMetProVal 220
Db 586 GTTGAAGAAATATCCCACTCCAACTCCAAAGCTGATACGAAACCAATCATGATCCAGTC 645
Oy 221 GluAspValPheSerIleAlaGlyArgGlyThrValValIleThrGlyArgIleGluArgLy 240
Db 646 GAAGACGATTCATCATCACTGGACGTGGTACTGTGCTACAGGCGCGTGGAAACGTGCT 705
Oy 241 ValValLybValGlyAspGluValGluIleValGlyIleArg---ProThrGluLybThr 255
Db 706 GAAGTTCGCGGTGGTGAACGAAGCTTAAATCGTTGATATTAAGACGAACACTCTTAACA 765
Oy 260 ThrValThrGlyValGluMetPheArgLybGluLeuGluLybGlyGluAlaGlyAspAsn 275
Db 766 ACTGTTACAGCTTGGAATGTTCCGTAAATTTATTAAGCTACAGCTGAACGACGGACAC 825
Oy 280 ValGlyValLeuLeuArgGlyThrLybLybGluGluValGluArgGlyMetValLeuCys 295

```
Db      826 ATCGGCTTTATTACGTGGTGTAGACAGTGAAGATATCGAACGTGACCAAGTATTAGCT 885
Qy      300 LysProGlySerIleThrProHisIleLysPheGluGlyGluIleTyrValLeuSerLys 319
      886 AAACCAAGCTACATCATCTCCACACAAATTCAAAGCTGAAGTATACGTATTATCAAAA 945
Qy      320 GluGluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArg 339
      946 GAAGAAGGGGAGCTGCACACTCCATTCTTCACTAAGTACCGTCCCTCAATTCTACTTCGGT 1005
Db      340 ThrThrAspValThrGlySerIleThrLeuPProGluGlyValGluMetValMetProGly 359
      1006 ACAACAGACGTTACTGGTGTGTAGAAATTGCCAGAGGTACTGAATGGTAATGCCCTGGT 1065
Qy      360 AspAsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLys 379
      1066 GATTACGTTGCTATGAGCGTTGAATTAAATTCAACCAATGCTTTGGAAGACGGAACTCGT 1125
Qy      380 PheAlaIleArgGluGlyGlyArgThrValGlyAlaGlyValIleSerAsnIleIleGlu 399
      1126 TTCTCTATTGCGAAGGGGAGTACTGTAGGTTCAGGCGTTGTACTGAATCGTTAAA 1185
```

Search completed: April 24, 2005, 06:16:41

Job time : 917.259 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2005, 05:33:37 ; Search time 229.225 Seconds

(without alignments)
2848.182 Million cell updates/sec

Title: US-10-039-183A-4

Perfect score: 2043
Sequence: 1 MAKEFNRTKPHVNIIGTIGH.....FAIRGKRTGAGVSNIE 399

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO/seq01/US10039183/runat_22042005_122426_20141/app_query.fasta_1.1038
-DB=Issued Patents NA -QFMT=fastext -SUFFIX=eml -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10039183 @cgn_1_1_219 @runat_22042005_122426_20141 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DISPATCH=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/pdata/1/1na/5A.COMB.seq:
2: /cgn2_6/pdata/1/1na/5B.COMB.seq:
3: /cgn2_6/pdata/1/1na/6A.COMB.seq:
4: /cgn2_6/pdata/1/1na/6B.COMB.seq:
5: /cgn2_6/pdata/1/1na/ECTUS.COMB.seq:
6: /cgn2_6/pdata/1/1na/Backfile1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1585.5	77.6	1191	4	US-09-902-540-8380
2	1585.5	77.6	7035	4	US-09-902-540-878
3	1579.5	77.3	1185	4	US-08-743-637B-185
4	1579.5	77.3	1830121	4	US-09-557-884-1
5	1579.5	77.3	1830121	4	US-09-557-884-1
6	1579.5	77.3	1830121	4	US-09-643-990A-1
7	1579.5	77.3	1830121	4	US-09-643-990A-1
8	1572.5	77.0	1245	3	US-09-134-001C-2540
9	1571.5	76.9	1185	3	US-09-218-197-1
10	1571.5	76.9	15598	4	US-08-956-171E-82
11	1571.5	76.9	15598	4	US-08-781-986A-82
12	1570.5	76.9	49617	4	US-09-596-002-28

13	1555.5	76.1	1185	4	US-09-492-709A-86	Sequence 86, Appl
14	1555.5	76.1	1254	4	US-09-489-039A-3648	Sequence 3648, Ap
15	1555.5	76.1	1260	4	US-09-489-039A-3750	Sequence 3750, Ap
16	1551.5	75.9	640681	4	US-09-790-988-1	Sequence 1, Appl
17	1525	74.6	1230	4	US-09-252-991A-4767	Sequence 4767, Ap
18	1525	74.6	1356	4	US-09-252-991A-4775	Sequence 4775, Ap
19	1521	74.4	1224	4	US-09-252-991A-4775	Sequence 4775, Ap
20	1521	74.4	1230	4	US-09-252-991A-4733	Sequence 4733, Ap
21	1513.5	74.1	1197	4	US-09-583-110-747	Sequence 747, Ap
22	1506	73.7	1122	4	US-09-134-000C-1261	Sequence 1261, Ap
23	1504	73.6	1191	4	US-09-107-532A-379	Sequence 1379, Ap
24	1500	73.4	1131	4	US-09-107-532A-979	Sequence 979, App
25	1465.5	71.7	4411529	3	US-09-103-840A-1	Sequence 1, Appl
26	1452.5	71.1	4403765	3	US-09-103-840A-2	Sequence 2, Appl
27	1428.5	69.9	1230025	4	US-09-198-452A-1	Sequence 1, Appl
28	1427.5	69.9	1230230	4	US-09-458-185A-1	Sequence 1, Appl
29	1413.5	69.2	1155	4	US-09-328-352-1547	Sequence 1547, Ap
30	1403.5	68.7	2996	3	US-08-961-527-260	Sequence 260, App
31	1395.5	68.3	580073	4	US-08-545-528D-1	Sequence 1311, Ap
32	1382.5	67.7	1122	4	US-09-540-236-22	Sequence 22, Appl
33	1311.5	64.2	1113	4	US-09-489-039A-4624	Sequence 4624, Ap
34	1266.5	60.0	1554	4	US-09-949-016-2750	Sequence 2750, Ap
35	1205.5	59.0	929	4	US-09-710-279-1311	Sequence 1311, Ap
36	1205.5	59.0	2997	4	US-09-710-279-3723	Sequence 3723, Ap
37	1065.5	52.2	1230	3	US-09-140-466-1	Sequence 1, Appl
38	940	46.0	864	4	US-09-107-433-1775	Sequence 1775, Ap
39	905.5	44.3	2592	3	US-09-221-017B-642	Sequence 642, App
40	887.5	43.4	3173	4	US-09-581-822-7	Sequence 7, Appl
41	829.5	40.6	690	4	US-09-543-681A-3724	Sequence 2724, Ap
42	792.5	38.8	3719	1	US-08-920-827-10	Sequence 10, Appl
43	792.5	38.8	3719	1	US-08-920-827-10	Sequence 10, Appl
44	792.5	38.8	3719	1	US-08-921-177-10	Sequence 10, Appl
45	792.5	38.8	3719	1	US-08-362-577C-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-902-540-8380
; Sequence 8380, Application US/09902540
; Patent No. 6833447

GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: MYXOCOCCUS XANTHUS Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 8380
LENGTH: 1191
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-8380

Alignment Scores:

Pred. No.: 2,126-171
Score: 1585.50
Percent Similarity: 86.50%
Best Local Similarity: 75.75%
Query Match: 77.61%
DB: 4
Gaps: 2

US-10-039-183A-4 (1-399) x US-09-902-540-8380 (1-1191)

QY 1 MetAlaLygJULyPheAsnArGrThLyPProHIsVaIaenILeGlyThrILleGlyHis 20
DB 1 ATGCCCAAGAGAAAGTTGAGAGGTAAACAAGCCCAAGTGAACATCGGACGATCGACAC 60

QY 21 ValAspHisgLyLysThrThrLeuSerAlaAlaIleSerAlaValIleuSerLeuysgLy 40
Db 61 GTGACCAACGGCAAGCGTCCGTGAACGGCCGACATCAACCAAGGTCTGGCGAAGACGGGC 120
QY 41 LeuAlaGluMetLysAspTyrTrpAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
Db 121 GGCGCCACGCTTCTCGGCGGACGCTGAATGACAAAGGCCCGGAGAGGAGTGAAGCGCGGT 180
QY 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAlaAspArgHisTyrAlaHisVal 80
Db 181 ATCAAGATTTCACCTCGACGCTGAAGATGCCAGAGTCCGACCGGACCTACCGCCACGTC 240
QY 81 AspCysProGluHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGluMetAsp 100
Db 241 GACTGTCCGGGCGACGGCCGACTACGGAAGAACATGATCAACGGGCGCGCGCAGATGAC 300
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgLysIle 120
Db 301 GGCGCATCTCGTGGTGTGCGGCGGAGCGGCCGATGCCAGACGCGTGAACATC 360
QY 121 LeuLeuSerArgGluValGlyValProHisIleValIlePheLeuAsnLysGluAspMet 140
Db 361 CTGCTGGCGCGCCGACAGTCCGCTTCCGTACATCGTGTCTTCTGAACAGGATGACATG 420
QY 141 ValAspAspGluGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
Db 421 CTGACGACGCCCGACGCTGGCGAGCTCGTGAAGATGAAATGCCGACCTCGTGAAGAG 480
QY 161 TyrGluPheProGluLysAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 481 TACGAGTTCCTCGGAGATGACATCCCATCATCCCGGCTCGGCGCTGAAGCGCTGAG 540
QY 181 GluAlaLysAlaGlyAsnValGlyGluTyrGlyGluLysValValLeuLysLeuMetAla 199
Db 541 -----GGTACACACGACGACATCGCGGACCGGCGCATCTCTGAAGCTGATGAG 588
QY 200 GluValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetPro 219
Db 589 GCGGTGGACAGCTACATCCGACGCGCGGCGGACGGAAGGCCCTTCTGATGCGCG 648
QY 220 ValGluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArg 239
Db 649 GTGAGGAGCGTTCCTCATCTCCGCGCGCGGACGCGTGGCCACGCGCGCTCGACGCGC 708
QY 240 GlyValAlaLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThr 259
Db 709 GGTATCATCAAGAGTGGCGGAGAGTGAAGTCTGTTGCTGCGCCGACGACGAGAGAGC 768
QY 260 ThrValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsn 279
Db 769 GTCGTCACGGGCGTGGAGATGTTCCGCAAGCTGTGAGACCAAGGCGATGGCGGCGACAC 828
QY 280 ValGlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCys 299
Db 829 ATCGCGCGCGTGGCGCGCTGGAAGCGCGAGACATGAGCGCGCGCGATGCTGGCGC 888
QY 300 LysProGlySerIleThrProHisLysPheGluGluIleTyrValLeuSerLys 319
Db 889 AAGCGCGGACGATCAACCGGACACCAAGTTCMAAGCGCACATCTACGCTGTCGAGAG 948
QY 320 GluGluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGluPheTyrValArg 339
Db 949 GAAGAGGGCGGTGTCACACCGCGTTCCTCAAGGGGTACCGCCGAGTTCATCTTCGCG 1008
QY 340 ThrThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGly 359
Db 1009 ACCACGAGCGTACCGGCTCGGTGAAGCTGCCGAGAGAGTGAATGTGTATGCGCGGC 1068
QY 360 AspAsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuLysThrLys 379
Db 1069 GACACATCGCATCGAGGTGAAGTCTCATACCGCGGTGGCGTGAAGAGAGAGTGGCGC 1128
QY 380 PheAlaIleArgGluGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu 399

Db 1129 TTCGCTGTTCGCGAGGGGTGGCGCACCGTGGGCGCGCGCTGTGGCGAATCATTCGAG 1188
RESULT 2
US-09-902-540-878/C
; Sequence 878, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIORITY FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIORITY FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 878
; LENGTH: 7035
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) .. (7035)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-878
Alignment Scores:
Pred. No.: 3,28e-170 Length: 7035
Score: 1585.50 Matches: 303
Percent Similarity: 86.50% Conservative: 43
Best Local Similarity: 75.75% Mismatches: 49
Query Match: 77.61% Indels: 5
Gaps: 2
US-10-039-183a-4 (1-399) x US-09-902-540-878 (1-7035)
QY 1 MetAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
Db 3263 ATGGCCAAAGAAATTGACGCTGAACAGCCCACTGAACATGCGACATCGGACATC 3204
QY 21 ValAspHisgLyLysThrThrLeuSerAlaAlaIleSerAlaValIleuSerLeuysgLy 40
Db 3203 GTGACCAACGGCAAGCGTCCGTGAACGGCCGACATCAACCAAGGTCTGGCGAAGACGGGC 3144
QY 41 LeuAlaGluMetLysAspTyrTrpAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
Db 3143 GGCGCCACGCTTCTCGGCGGACGCTGAATGACAAAGGCCCGGAGAGGAGTGAAGCGCGGT 3084
QY 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAlaAspArgHisTyrAlaHisVal 80
Db 3083 ATCAAGATTTCACCTCGACGCTGAAGATGCCAGAGTCCGACCGGACCTACCGCCACGTC 3024
QY 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGluMetAsp 100
Db 3023 GACTGTCCGGGCGACGGCCGACTACGTAAGAACATGATCAACGGGCGCGCGCAGATGAC 2964
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgLysIle 120
Db 2963 GGCGCATCTCGTGGTGTGCGGCGGAGCGGCCGATGCCAGACCGCGTGAACATC 2904
QY 121 LeuLeuSerArgGluValGlyValProHisIleValIlePheLeuAsnLysGluAspMet 140
Db 2903 CTGCTGGCGCGCCGACAGTCCGCTTCCGTACATCGTGTCTTCTGAACAGGATGACATG 2844
QY 141 ValAspAspGluGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
Db 2843 CTGACGACGCCCGACGCTGGCGAGCTCGTGAAGATGAAATGCCGACCTCGTGAAGAG 2784
QY 161 TyrGluPheProGluLysAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180

Db 2783 TACGAGTTCCTGGGATGATCATCCCATCATCCGGGCTCGGCTGAGGCGCTGGAG 2724
Qy 161 GUAUAlaValaGlyAsnValaGlyGluTgylgylu---LysValleuLysleuMetAla 199
Db 2723 -----GGTGAACACCGACCGACATCGCGAGCCGCCCATCTGTAAGCTGATGGAG 2676
Qy 200 GUUAlaAspAlaTyrIleProThrProGluTyrAspThrGluLysThrPheLeuMetPro 219
Db 2675 GCGGTGACAGCTACATCCCGACCGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2616
Qy 220 ValGluAspValPheSerIleAlaGlyArgGlyThrValValaThrGlyArgIleGluArg 239
Db 2615 GTGGAGGAGCTGTTCTCCATCTCCGCGCGCGGACGAGTGGCCACGCGCGCTCGAGCGCC 2556
Qy 240 GlyValValLysValaGlyAspGluValaGluIleValaGlyIleArgProThrGlnLysThr 259
Db 2555 GGTATCATCATAGTGGCGAGGAGGAGGAGGAGTGGTGGTGGTGGCGCGCGGAGGAGGAG 2496
Qy 260 ThrValThrGlyValaGluMetPheArgGlyGluLeuGluValaGlyValaGlyAspAsn 279
Db 2495 GTGCTGACGGGCGGTGAGATGTTCCGCAAGCTGTCGACACGAGGAGGAGGAGGAGGAG 2436
Qy 280 ValGlyValLysLeuArgGlyThrLysLysGluGluValaGluArgGlyMetValleuLys 299
Db 2435 ATCGCGCGGCTGCTGCGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2376
Qy 300 LysProGlySerIleThrProIleLysPheGluGluGluIleTyrValleuSerLys 319
Db 2375 AAGCGGCGAGATACCCCGACACCAAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2316
Qy 320 GluGluGlyGlyArgHisThrPhePhePhePhePhePhePhePhePhePhePhePhe 339
Db 2315 GAGAGGCGCGGCTGCTGACACCCGCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2256
Qy 340 ThrThrAspValThrGlySerIleThrLeuProGluGluValaGluMetValleuProGly 359
Db 2255 ACCAGGAGCGTGAACGGCTCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2196
Qy 360 AspAsnValLysIleThrValaGluLeuIleSerProValAlaLeuGluLeuGlyThrLys 379
Db 2195 GACAACTATCGCATGAGGTGAGCTCATCACCCCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 2136
Qy 380 PheAlaIleArgGluGlyGlyArgThrValaGlyAlaGlyValaIleSerAsnIleIleGlu 399
Db 2135 TTCGCTGTTCCGAGGAGGTGGCCGACCGTGGGCGCGGCGGCTGTCGCGGAGATCATCGAG 2076

RESULT 3
US-08-743-637B-185
; Sequence 185, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/743, 637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526, 840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35, 433
REFERENCE/DOCKET NUMBER: 850586, 90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5591
TELEFAX: (414) 277-5000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 1185 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
US-08-743-637B-185
ORGANISM: Haemophilus influenzae

Alignment Scores:
Pred. No.: 1,026-170 Length: 1185
Score: 1579.50 Matches: 297
Percent Similarity: 85.71% Conservative: 45
Best Local Similarity: 74.44% Mismatches: 52
Query Match: 77.31% Indels: 5
DB: 2 Gaps: 1

US-10-039-183a-4 (1-399) x US-08-743-637B-185 (1-1185)

Qy 1 MetaLysGlyLysPheAsnArgThrLysProHisValaAsnIleGlyThrIleGlyHis 20
Db 1 ATGCTTAAGAAATAATTGAACGTACAAACCGACGTAAAGCTGGTATCAATCGGCAC 60
Qy 21 ValAspHisGlyLysThrThrLeuSerAlaAlaIleSerAlaValleuSerLeuLysGly 40
Db 61 GTTGACCAAGGTAAACAACTTAAACACAGCAATCAACACCGTATGGCAAAACATTAC 120
Qy 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
Db 121 GGTGTGACAGGCGGTGACATTGACCAATTGATTAACGCCGCAAGAAAGCCGCTGCT 180
Qy 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
Db 181 ATTACCATCAACACTTACACAGTGAATACGATACACGAGCTCGCCACTATGACACGTA 240
Qy 81 AspCysProGlyHisAlaAspTyrValaLysAsnMetIleThrGlyAlaAlaGluMetAsp 100
Db 241 GACTGTCCGGACACCGCGACTATGTTAAATAATGATTACGTGGCGGACAAATGGAT 300
Qy 101 GlyAlaIleLeuValaIleSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
Db 301 GGTGCTATTATTAAGTACGACACACAGATGCTCTTAAGCCGCAATCTGTGAACATC 360
Qy 121 LeuLeuSerArgIleValaGlyProHisIleValaIlePheLeuAsnLysGluAspMet 140
Db 361 TTATTAGTGGCAAGTAGTGTTCATCATCATCATCTTAAACAAATGCGACATG 420
Qy 141 ValAspAspGluGluLeuGluGluValaGluMetGluValaArgGluLeuLeuSerAla 160
Db 421 GTAGATGACGAAGATTTAGTAATTAATCAAAATGAAAGTTCGTAACCTTACTCA 480
Qy 161 TyrGluPheProGlyAspAspThrProIleValaAlaGlySerAlaLeuArgIleGlu 180
Db 481 TATGACTTCCAGGTGACGATACACCAATCGTACGTGTTACGATTCACACGCTTA--- 537
Qy 181 GUAUAlaValaGlyAsnValaGlyGluTgylgyluLysValleuLysleuMetAlaGlu 200
Db 538 -----AACGGGTGACGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 585

201 ValaAspAlaTyrIleProThrProGluArgAspThrGluValThrPheLeuMetProVal 220
586 TTAGTACTTACATCCAGAACACAGACGTGGATTGACCAACCGTCTCTTCCAAATC 645
221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
646 GAAGATGTTCTCATCTCAGGTGCTGCTAGTAAACAGGTCTGTAGAACACAGGT 705
241 ValValValValGlyValAspGluValGluIleValGlyIleArgProThrGlnThrThr 260
706 ATTATCCGACACGGATGAGTAGAATCGTGGATATCAAGATACACGAAACTACT 765
261 ValThrGlyValGluMetPheArgGlyGluLeuGlyGlyValAlaGlyAspAsnVal 280
766 GTAAAGGGGTGTAATGTTCCGTAATATCTTGAAGAGCGTGCACAGTGAACATC 825
281 GlyValLeuLeuArgGlyThrIlySlyGluGluValGluArgGlyMetValLeuCysIlyS 300
826 GGTGATTATTCGTGTACCAAGCGTAAGAAATCGTGCACAGTATTAAGCGAA 885
301 ProGlySerIleThrProHisIlySlyPheGluGlyGluIleThrValLeuSerIlySgu 320
886 CCAAGTTCATACACACACACTGACTTCAATCAGAAAGTACGATTAATCAAAAGAT 945
321 GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArgThr 340
946 GAAGGTGGTGCATACATCTTCTTCAAGGTTACCGTCCCAANTTCATTTCCGTACA 1005
341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
1006 ACAGACGTGACTGTGATCATCAATTAACAGAAAGGCGTGAATGCTAATGCACGCGAT 1065
361 AsnValIlySleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrIlySph 380
1066 AACATCAAGATGACAGTAACTTAATCCACCAATTCGATGACAGCAAGTTAGCTTTC 1125
381 AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValIleSerAsnIleIleGlu 399
1126 GCAATCCGTAAGGTGGCGGTACAGTACGTGCGGTGTTGCCAAATCATCTCAA 1182

RESULT 4
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS V6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB18693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Alignment Scores:
Pred. No.: 8,496-166 Length: 1830121
Score: 1579.50 Matches: 297
Percent Similarity: 85.71% Conservative: 45
Best Local Similarity: 74.44% Mismatches: 52
Query Match: 77.31% Indels: 5
Gaps: 1
US-10-039-183a-4 (1-399) x US-09-557-884-1 (1-1830121)
1 MetAlaIlySguIlyPheAsnArgThrIlySProHisValAsnIleGlyThrIleGlyHis 20
671167 ATGTCTAAAGAAATTTGAACGTACAAACCGCACGTAAACGTGGTACATCGGCAC 671226
21 ValaAspHisGlyIlySThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuIlySgly 40
671227 GTTGACCAAGTAAACACTTAAACAGCGAATACACACCGTATTTGGCAAAACATTAC 671286
41 LeuAlaGluMetIlySAspTyrAspAsnIleAspAsnAlaProGluGluIlySguIlyArgGly 60
671287 GGTGTGACAGGCGCGTCAATTCGACCAATTTGTAACGCCAGAAAGAAACCGGTGGT 671346
61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
671347 ATTACATCAACACACTTCACAGTTGAATACGATACACCGACTCGCTATGACACGTA 671406
81 AspCysProGlyHisAlaAspTyrValIlySAsnMetIleThrGlyAlaAlaGluMetAsp 100
671407 GACTGTCCGGACACGCGCATGTATTAATAATGATTAATCTGTGGCGCACAAATCGAT 671466
101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
671467 GGTGCTATTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 671526
121 LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnIlySguIlyMet 140
671527 TTATTAGGTCCGCAAGTGGTGTTCATCATCATCTTAAACAAATGCGACATG 671586
141 ValaAspAspGlnGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
671587 GTAGATGACGAAGATTAATTAAGATTAGTCGAAATGGAAGTTCGTAACCTTATCTCAA 671646
161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
671647 TATGACTTCCAGCGTGAAGATACCAATCGTACGTGTTCAAGTCGTTA--- 671703
181 GluAlaIlySAlaGlyAsnValGlyIlyTrpGlyGlyValLeuIlySLeuMetAlaGlu 200
671704 -----AACGGGTGACGAATGCGGAAGAAATCTTGAAGTGAACAAACAC 671751
201 ValaAspAlaTyrIleProThrProGluArgAspThrGluValThrPheLeuMetProVal 220
671752 TTAGTACTTACATCCAGAACACAGACGTGGATTGACCAACCGTCTCTTCCAAATC 671811
221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
671812 GAAGATGTTCTCATCTCAGGTGCTGCTAGTAAACAGGTCTGTAGAACACAGGT 671871
241 ValValValValGlyValAspGluValGluIleValGlyIleArgProThrGlnThrThr 260


```

Db      671872 ATTATCCGTACAGTGTAGTACAGTAAATCGTCGGTATCAAGATACACGAAAACTACT 671931
Qy      261 ValThrGlyValIGlUmePheArGlySGlUleuGluVysGlyGluAlaGlyAspAsnVal 280
Db      671932 GTTACGGGGTGTGAATGTTCCGTAAATTACTTGACGAAGGTCGTGCAGGTGAAGAACATC 671991
Qy      281 GlyValLeuLeuArGlyThrLylySGlUglUValIGluArGlyUmeValLeuCylys 300
Db      671992 GGTGCATTATTACGTGTACCAACGTGAAGAAATCGAACGTGTCAATATTAGCGAAA 672051
Qy      301 ProGlySerIleThrProHisLylySphGluGlyIleTyrrAlaLeuSerLylyGlu 320
Db      672052 CCAGGTTCAATACACACACACCTGCTTCGAATACGAAGTGTACGATTATTAACAAAGAT 672111
Qy      321 GluGlyGlyValArgHisThrProPheThrIleAsnTyrrArgProGlnPheTyrrAlaArgThr 340
Db      672112 GAAGGTGTGTCATCAATCTCATCTTTCGAAGGTTACCGTCACCAATTCTAATTCGTGACA 672171
Qy      341 ThrAspValThrGlySerIleThrLeuProGluGlyValIGlUmeValUmeProGlyAsp 360
Db      672172 ACAGACGTGACTGTGACATCAATCAATTACAGAAAGCGGTGAATGTATGCCAGCGCAT 672231
Qy      361 AsnValLylyIleThrValIGlUleuIleSerProValAlaLeuGluLeuGlyThrLySph 380
Db      672232 AACATCAAGATGACAGTAACTTAATCCACCAATTGCGATGAGACCAAGTTTACGTTTC 672291
Qy      381 AlaIleArGlyGlyValArgThrValIGlyAlaGlyValValSerAsnIleIleGlu 399
Db      672292 GCAATCCGTGAAGGTGGCGTACAGTACAGTGTGGCGGTGTTCGCAAAAATCATCAAA 672348

RESULT 5
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS V6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```

```

US-09-557-884-1
Alignment Scores:
Pred. No.: 8,496-166 Length: 1830121
Score: 1579.50 Matches: 297
Percent Similarity: 85.71% Conservative: 45
Best Local Similarity: 74.44% Mismatches: 52
Query Match: 77.31% Indels: 5
DB: 4 Gaps: 1

US-10-039-183a-4 (1-399) x US-09-557-884-1 (1-1830121)
Qy      1 MetAlaLySGlUlyPheAsnArGThrLySProHisValAsnIleGlyThrIleGlyHis 20
Db      597082 ATGTCTAAAGAAAATTTCAGACGTACAAAACCGACGTAAAGTGGTTCATTCGGCCAC 597023
Qy      21 ValAspHisGlyLySThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLySgly 40
Db      597022 GTTGACCAACGGTAAACAACTTTAACAGACGAAATCAACACCGTATTAGCAAAACACTAC 596963
Qy      41 LeuAlaGluMetLyAspTyrrAspAsnIleAspAsnAlaProGluGluGlyValArgGly 60
Db      596962 GGTGTGTGACGCGCGTGTGATTCACCAATGTATACCGCGCAGAAAGAAACCGGTGTGT 596903
Qy      61 IleThrIleAlaThrSerHisIleGlyIleTyrrGluThrGluAsnArGHisTyrrAlaHisVal 80
Db      596902 ATTACCATCAACACCTTCACACGTGAATACATACCAACCGACTCGCCACTACGACACGT 596843
Qy      81 AspCySProGlyHisAlaAspTyrrValLySAsnMetIleThrGlyAlaAlaGluMetAsp 100
Db      596842 GACTGTCCGGGACACGCGGACTATGTAAATAATATGATTACTGTGCGGCAAAATGAT 596783
Qy      101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
Db      596782 GGTGTATTATTAGTGTACGACACCAACAGTGTCTTATGCCCAAACTCGTAAACATAC 596723
Qy      121 LeuLeuSerArGSGluValIGlyValProHisIleValAlaPheLeuAsnLySGluAspMet 140
Db      596722 TTATTAGTCCGCCAAGTACGTGTCTTCATCATCATCATCTTTTAAACAAATGCCACATG 596663
Qy      141 ValAspAspGluGluLeuLeuGluLeuValIGlUmeGlyValArgGluLeuLeuSerAla 160
Db      596662 GTAGATGACGAAGAATTATTAGAAATTAGTCGAATAGGAAGTTCGTAATCTTCACTCA 596603
Qy      161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArGAlaLeuGlu 180
Db      596602 TATGACTTCCCGACGTGACGATACCAACATCGTACGTGTTCAGCATTCACACGCTTA--- 596546
Qy      181 GluAlaLySAlaGlyAsnValIGlyIleTyrrGlyGlyLySValLeuLyLeuMetAlaGlu 200
Db      596545 -----ACCGCGTACGAAATGGGAAGAAATACTTGAGTTACCAAAACAC 596498
Qy      201 ValAspAlaTyrrIleProThrProGluArGAspThrGlyLySThrPheLeuMetProVal 220
Db      596497 TTAGTACTTTCATCTCCAGAACCAACAGTGTGATGACCAACCGTCTTCTTCCAAATC 596438
Qy      221 GluAspValPheSerIleAlaGlyArGlyThrValValThrGlyArGlyIleGluArGly 240
Db      596437 GAAGATGTGTTCTCAATCTCAGTGTGTGTCTGTAGAACAGGTGTGTGTACAAAGAGGT 596378
Qy      241 ValValLySValIGlyAspGlyValIGlyIleValIGlyIleAspProThrGlnLySThrThr 260
Db      596377 ATTATCCGTACAGGTGATGAATGAATTCGTCGTATCAAGATACAGCGAAACTACT 596318
Qy      261 ValThrGlyValIGlUmePheArGlySGlUleuGlySGlyGluAlaGlyAspAsnVal 280
Db      596317 GTTACGGGGTGTGAATGTTCCGTAAATTACTTGACGAAGGTCGTGCAGGTGAAGAACATC 596258
Qy      281 GlyValLeuLeuArGlyThrLylySGlUglUValIGluArGlyUmeValLeuCylys 300
Db      596257 GGTGCATTATTACGTGTACCAACGTGAAGAAATCGAACGTGTCAATATTAGCGAAA 596198
Qy      301 ProGlySerIleThrProHisLylySphGluGlyIleTyrrAlaLeuSerLylyGlu 320

```

|||||
Db 596197 CCAGGTTCAATCACACACACTGACTTGAATCGAAGGTAGCGTATTATCAAAAGAT 596138
|||
Qy 321 GUGUGYAGYAGHietHrProphetherAntyYrYrProGlnPheYrValArgThr 340
|||
Db 596137 GAAAGGTGGTCGTCATCTCATCTTCCAAAGGTTACCGGCCAACATTTCTATTCCGTACA 596078
|||
Qy 341 ThrAPValThrGlySerIleThrLeuProGlnGlyValGlnMetValMetProGlyAsp 360
|||
Db 596077 ACAGACGTACCTGTCATCATCTTCAATTCACAGAGCGGTGAAATGTGTAATGCCAGCGCAT 596018
|||
Qy 361 AsnValIleYrIleThrValGlnLeuIleSerProValAlaLeuGlnLeuGlyThrIysPhe 380
|||
Db 596017 AACATCAAGATGACAGTAAGCTTAATCCACCAATGCGATGATCAAGTTAGCGTTTC 595958
|||
Qy 381 AlaIleArgGlnGlyYrYrGhrThrValGlnValGlnValSerAsnIleIleGlu 399
|||
Db 595957 GCAATCCGTGAAGGTGGCCGTACAGTAGTGACGGCGTTTGCAGAAATCATCAAA 595901
|||
RESULT 6
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS V6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Alignment Scores: 8,49e-166 Length: 1830121
Pred. No.: 1579.50 Matches: 297

Percent Similarity: 85.71% Conservative: 45
Best Local Similarity: 74.44% Identities: 52
Query Match: 77.31% Mismatches: 5
DB: 4 Gaps: 1
US-10-039-183A-4 (1-399) x US-09-643-990A-1 (1-1830121)
Qy 1 MetAlaIySGlUlySPheAsnAArgThrIysProHisValAsnIleGlyThrIleGlyHis 20
|||
Db 671167 ATGTCTAAAGAAAATTGTAAGCTACCAACCGACGTAACGTGGTACATCGGCAC 671226
|||
Qy 21 ValAspHisGlyYrYrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuIysGly 40
|||
Db 671227 GTTGACCGACGGTAAGAACTTTAAACACAGCAATCAACACCGTATGGCAAAACATTAC 671286
|||
Qy 41 LeuAlaGlnMetIysAspThrAspAsnIleAspAsnAlaProGlnGlnIysGlnArgIy 60
|||
Db 671287 GGTGGTGCAGCGCGGTCATTCGACCAATGTATACCGCCGAGAAAGAAACCGGTGGT 671346
|||
Qy 61 IleThrIleAlaThrSerHisIleGlnIyrglnThrglnAsnAArgIleYrAlaHisVal 80
|||
Db 671347 ATTACCATCAACACTTTCACACGTTGATACGATACACCGACTCGCACTATGCACAGTA 671406
|||
Qy 81 AspCysProGlnYHisAlaAspIYrValIysAsnMetIleThrGlyAlaAlaGlnMetAsp 100
|||
Db 671407 GACTGTCCGGGACACGCCGACATGTATAAATATGATTACTGTGTCGGCGCACAAATGAT 671466
|||
Qy 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGlnHisIle 120
|||
Db 671467 GGTGCTATTATTAGTATGTAAGACAGCAACAGATGGTCTTATCCGCAACTCGTAACACATC 671526
|||
Qy 121 LeuLeuSerArgGlnValGlyValProHisIleValAlaPheLeuAsnIysGlnAspMet 140
|||
Db 671527 TTATTAGCTCCGCAAGTAGTGGTTCATACATCATCGTATCTTAAACAAATGCGACATG 671586
|||
Qy 141 ValAspAspGlnGlnLeuLeuGlnLeuValGlnMetGlnValArgGlnLeuLeuSerAla 160
|||
Db 671587 GTAGATGACGAAAGATTATGAAATTAAGTACGAAAGAAAGTTCGAACTTCTACTCNA 671646
|||
Qy 161 TyrGlnPheProGlnYAspAspThrProIleValAlaGlySerAlaLeuAsnArgAlaGln 180
|||
Db 671647 TATGACTTCCAGGTGACGATACCAATCGTACGTGTTCGACTTACACACGCTTA--- 671703
|||
Qy 181 GlnAlaIyAlaGlyAsnValGlyIyThrProGlnIyGlnIyValLeuIyLeuMetAlaGln 200
|||
Db 671704 -----AACGGGTGACGAATGGAGAAAGAAATCTTGAGTTGCAAAACAC 671751
|||
Qy 201 ValAspAlaIYrIleProThrProGlnArgAspThrglnIyYrThrPheLeuMetProVal 220
|||
Db 671752 TTAGATTAATTACATCCAGAACCAAGACGTGATGACCAACCGTCTCTTCCAAATC 671811
|||
Qy 221 GlnAspValPheSerIleAlaGlyYrArgIyThrValIleThrglyYrAlaIleGlnArgIy 240
|||
Db 671812 GAAAGATGCTCTCAATCTCAGGTGCTGCTGTAGTAACAGGTGCTGTAGAACAGAGT 671871
|||
Qy 241 ValValIySValGlyAspGlyValGlnIleValGlyIleArgProThrglnIySerThrThr 260
|||
Db 671872 ATTATCCGTACAGGTGATGAAGTAGAATCGCGTATCAAAAGATACGCAAACTACT 671931
|||
Qy 261 ValThrglyValGlnMetPheArgIyGlnLeuGlnIySGlyValAlaGlyAspAsnVal 280
|||
Db 671932 GTAACGGGTGTTGAATGTTCCGTAAATTACTTTCGAAAGGTGCGTGAAGAAACATC 671991
|||
Qy 281 GlyValIleLeuAsnArgIyThrIyIySGlnValAlaGlnArgIyMetValLeuIySlys 300
|||
Db 671992 GGTGCATTATTAATCGGTACCAAGGTGAAGAAATCGACGTGCAATTTATACGAAA 672051
|||
Qy 301 ProGlySerIleThrProHisIyIySPhesGlnGlnIyGlnIyYrValLeuSerIySgln 320
|||
Db 672052 CCAGGTTCAATCACACACACTGACTTCAATAGAAAGTACGTATTTATCAAAAGAT 672111
|||
Qy 321 GlnGlyYrYrGhrIeThrProphetherAntyYrYrProGlnPheYrValArgThr 340
|||


```

Oy      361  AenValIyVgIleThrValGlutLeuIleSerProValAlaLeuGlutLeuLyrHisPhe 380
          ::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      596017 AACATCAAGATGACGATGATGCTTAATCCACCCAAATTCAGATCAAGATTACGTTTC 595958
          ::::::::::::::::::::::::::::::::::::::::::::::::::::
Oy      381  AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValIValSerAsnIleIleGlu 399
          ::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      595957 GCATTCCTGGAAGCGTGGCCCTGACAGTAGTGCGACGCGTTGTCGAAAAATCATCAAA 595901
          ::::::::::::::::::::::::::::::::::::::::::::::::::::

RESULT 8
US-09-134-001C-2540
; Sequence 2540, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2540
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2540

```

Alignment Scores:	
Pred. No.:	6.55e-170
Score:	1572.50
Percent Similarity:	85.71%
Best Local Similarity:	74.44%
Query Match:	76.97%
DB:	3
US-10-039-183A-4 (1-399) x US-09-134-001C-2540 (1-1245)	
Length:	1245
Matches:	297
Conservative:	45
Mismatches:	52
Indels:	5
Gaps:	1

Qy	1	MetalAlayseGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis	20
Dp	61	ATGGCAAAAGAAAATTGGTCCTCAAAAGAACATCCAAATTTGGTACTATGGCTCAC	120
Qy	21	ValAspHisGlyLysThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLysGly	40
Dp	121	GTTGACCATGCTAAACCACTTAAACAGCTGGCTATGCAACTGTATTAGTAAAAATGGT	180
Qy	41	LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly	60
Dp	181	GACACTGGTGCACAACATCATACGATGATGATGACAAACGCTCCGAAGAAAAAGAACGTGGT	240
Qy	61	IleThrIleAlaThrSerHisIleGlyTyrGluThrGluAsnArgHisETyrAlaHisVal	80
Dp	241	ATTACATCATATCTGCACATACGAAATCCAAATCGAACAACGCTATATGCTCAGCTT	300
Qy	81	AspCysProGluHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGluMetAsp	100
Dp	301	GACTGCCCAAGAACCGCTGACTATGTTAAAAACATGATACATCGTGCAGCTCAATATGGAC	360
Qy	101	GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle	120
Dp	361	GGCGGATTCCTAGTGTATCTGCTGTGACGGTCCAAAGCCAAACCTCGGAACACATC	420
Qy	121	LeuLeuSerArgGluValGlyValProHisIleValAlaPheLeuAsnLysGluAspMet	140
Dp	421	TTATTTATCAAGTAAACGTGGTGTATCCAGCATTAAGTTGATCTTTAAACAAAGTTGACATG	480
Qy	141	ValAspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla	160
Dp	481	GTTGACCAACAAAGATTTATTGAAATTAGTTGAAATGGAAAGTTCGTGACTTTATTAAAGCGAA	540

QY	161	YrGluPheProGlyAAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu	180
Db	541	TATGCTTCCAGGAGGACGATGACTCGTAAAGCGGTGGTCTTGCAATTAAGCATTAAGAA	600
QY	181	GIuAlaIyBaIaGlyAsnValaGlyIuTrpGlyGluIyValIleuIyBseuMeIaGlu	200
Db	601	-----GGCGATGCTGAATATCGAAACAAAATCTTGAATTAATGCAAGA	645
QY	201	ValAspAlaTyrlIleProThrProGluIuArgAspThrGluIyThrPheIuMetProVal	220
Db	646	GTTGATGATTCATTCCTCAATCCAACTCCAAAGCGTGAATTCGACAAACCATTCAGAGCCAGTT	705
QY	221	GIuAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly	240
Db	706	GAGAGCGATATCTCAATCACTGCTGCTGGTACTGTGTGCTACAGGCCGTGTTGAACGTGGT	765
QY	241	ValValIyIyValGlyAspGluValGluIleValGlyIleArgProThrGluIyThrThr	260
Db	766	CAAACTCAAAAGTGGTGGAAGAGTTAAATCAACGGTATGACAGAACTTCTAAACAACT	825
QY	261	ValThrGlyValaGluMetPheArgIyGluLeuGluIyGlyAlaGlyAspAsnVal	280
Db	826	GTTACTGCTGTAAGAAATGTTCCGTAAATTAATTAAGACTACGCTGAAGCTGGTGAACAATC	885
QY	281	GIyValIleuLeuArgGlyThrIyIyGluGluValaGluArgGlyMetValIeuCyIyS	300
Db	886	GGTGGCTTAATTAACGGTGGTGGTGCACGTAAAGCGTACAAACGTCGCAAGTATTAAGCTGCT	945
QY	301	ProGlySerIleThrProHlybIySerPheGluGluIyGluIleTyrValIeuSerIyGlu	320
Db	946	CCTGGTCTTAATTAACAACAACAACAAATTAACAAAGCTGAAGTATGATTAATTAACGAT	1000
QY	321	GIuGlyGlyArgHlyeThrProPheBheThrAsnTyrrArgProGluPheTyValArgThr	340
Db	1006	GAAGGTGACGCAACACTCCATCTTCTCACTAATCTATCGCCCAATTCATATTCCGCTACT	1060
QY	341	ThrAspValThrGlySerIleThrIeuProGluGlyValaGluMetValIeuProGlyAsp	360
Db	1066	ACTGACGTAACTGCTGTTGTAACTTACACAGAAAGTACAGAAATGGTTATAGCTGGCGAC	1120
QY	361	AsnValIyIleThrValaGluIeuIleSerProValAlaIeuGluIeuGlyThrIyIyPhe	380
Db	1126	AACGTTGAATAGACAGTGAATTAATCGCTCAATCGATCGATCGAAGACGAACCTGTTTC	1180
QY	381	AlaIleArgGluGlyGlyArgThrValaGlyAlaGlyValaIeuSerAsnIleIleGlu	399
Db	1186	TCAATTCGTGAAGGTGAAGCATAGTGGATGACGCGTTGTAACGAAATCTTTGAA	1242

```

RESULT 9
US-09-218-197-1
; Sequence 1, Application US/09218197A
; Patent No. 6451556
; GENERAL INFORMATION:
; APPLICANT: Kallender, Howard
; APPLICANT: Van Horn, Stephanie
; TITLE OF INVENTION: EP-TU
; FILE REFERENCE: GM10185
; CURRENT APPLICATION NUMBER: US/09/218,197A
; CURRENT FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-218-197-1

Alignment Scores:
Pred. No.:      8,39e-170      Length:      1185
Score:          1571.50        Matches:      296
Percent Similarity: 86.47%    Conservative: 49
Best Local Similarity: 74.13% Mismatches:     49
Query Match:     76.93%      Indels:         5

```

DB: 3 Gaps: 1

US-10-039-183a-4 (1-399) x US-09-218-197-1 (1-1185)

QY 1 MetAlaLVeGluVpPheAsnArGThrLyPProHISValAsnIleGlyThrIleGlyHis 20

DB 1 AAGGAGAAAAGAAAATTCATGCTCTTAAAGAACATGCCAATATCGGTATCTTCGCTCAC 60

QY 21 ValAspHISGlyVLeThrThreUsEserAlaIleSeraIValIleUserLeuLySgIly 40

DB 61 GTTGACCATGTAAGAACACATTACAGACAAATCGCTACTGATATTACCAAAAAATGCT 120

QY 41 LeuAlaGluMeLVeAspTYrAspAsnIleAspAsnAlaProGluGluVpSgIly 60

DB 121 GACTCAGTTCACAAATCATATGATGATGACAAAGCTCCAGAGAAAAGAAAGAGCTGCT 180

QY 61 IletHrIleAlaThSerHISileGlyUTyrGluThrGluAsnArGHisTYrAlaHISVal 80

DB 181 ATCACAATCAATATCTTCACATGAGTCCAAACTGACAAACGTCACACCTCACGCTT 240

QY 81 AspCysProGluVHISAlaAspTYrValIleAspMetIleThrGlyAlaIleAlaGluMeLVe 100

DB 241 GACTCCCAAGACAGAGCTGACTACGTTAAACATGATCATCTGCTGCTCAAAATGAC 300

QY 101 GlyAlaIleLeuValIleSeraIAlaAspGlyProMetProGluThrArgSgIlyHisIle 120

DB 301 GCGCGTATCTTAGTAGTATCTGCTGCTGACGCTCCAAATGCCAACAACCTGTAACACAT 360

QY 121 LeuLeuSerArgGluValIleGlyValProHISileValPheLeuAsnLySgIlyAspMet 140

DB 361 CTTTATACACGTAAAGTGTGCTGCTACAGCATTAAGTATCTTAAACAAAGTTACACATG 420

QY 141 ValAspAspGluGluLeuGluGluMeLVeGluValIleArgGluLeuLeuSeraIle 160

DB 421 GTTGACGATGAAGAAATATTAGAAATTAAGAAATGAAAGTTCGTGATTAATTAAGCGGA 480

QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySeraIleLeuArgAlaLeuGlu 180

DB 481 TATGATCTCCAGAGAGATGATGATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

QY 181 GluAlaLVeAlaGlyAsnValIleGlyUTyrGlyGluVpSgIlyLeuLySgIlyMeLVe 200

DB 541 -----GCGCATGCTCAATACGAAGAAATCTTAGAAATTAATGAAGCT 585

QY 201 ValAspAlaTYrIleProThrProGluVpSgIlyThrValPheLeuMetProVal 220

DB 586 GTAGATCTTACATTCACATCCAGAACGTGATTCGACAAACCATTCATGATGACAGCT 645

QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValIleThrGlyArgIleGluArgGly 240

DB 646 GAGGACGTAATCTCAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705

QY 241 ValValIleValIleGlyAspGluValIleValIleGlyIleArgProThrGluLySgIlyThr 260

DB 706 CAATCAAAAGTTGAGAGAAAGTGAATCATCGGTTTACATGACACATCACTAAACAAC 765

QY 261 ValThrGlyValIleGluMeLVePheArgGlySgIlyLeuGluVpSgIlyAlaGlyAspVal 280

DB 766 GTTACAGGTTGAAATGTTCCGTAAATTAATTAAGCTAGCGTAAAGCTGCTGACAACT 825

QY 281 GlyValIleLeuValArgGlyThrLySgIlyGluValIleGluArgGlyMeLVeValIleUcys 300

DB 826 GGTGCATTTATTCGTTGCTGCTGTAAGACGTAACAGTGCAGATTAAGCTGCT 885

QY 301 ProGlySerIleThrProHISValLySgIlyPheGluGlyGlyIleTYrValIleUserLySgIly 320

DB 886 CTTGCTTCAATTAACACATGATTAAGCAAGATTAAGCAAGATTAAGCAAGATTAAGCAAG 945

QY 321 GluGluGlyValArgHISThrProPheThrAspTYrArgProGluPheTYrValArgThr 340

DB 946 GAAGGTGACGTCACATCTCTTCTAACTATGCTGCTCAAACTTCAATTCATTTCCGTA 1005

QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValIleGluMeLVeLVeMetProGlyAsp 360

DB 1006 ACTGACGTAAGTGTGCTTCTCACTTACCAAGAGTACTGAATGTAATGCTGCTGAT 1065

QY 361 AsnValIleIleThrValIleGluLeuIleSerProValAlaLeuGluLeuGlyThrLySgIly 380

DB 1066 AACGTTGAATGACGTAATTAATTCCTCAATCGGATGAAAGACGTAATCTGCTTTC 1125

QY 381 AlaIleArgGluGlyValArgThrValIleGlyAlaGlyValIleSeraIleIleGlu 399

DB 1126 TCAATCCCGAAGTGAAGTACTGTAAGTACGCGTTGTTACTGAATCATTA 1182

RESULT 10

US-08-956-171E-82

; Sequence 82, Application US/08956171E

; Patent No. 6593114

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

Gail H. Choi

Patrick S. Dillon

Craig A. Rosen

Steven C. Barash

Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 15598 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 82:

US-08-956-171E-82

Alignment Scores:

Pred. No.: 4,478-168 Length: 15598

Score: 1571.50 Matches: 296

Percent Similarity: 86.47% Conservative: 49

Best Local Similarity: 74.19% Mismatches: 49

Query Match: 76.92% Indels: 5

DB: 4 Gaps: 1

US-10-039-183a-4 (1-399) x US-08-956-171E-82 (1-15598)

QY 1 MetAlaLVeGluVpPheAsnArGThrLyPProHISValAsnIleGlyThrIleGlyHis 20

QY	81	AApCyPProGlyYH1SA1LAaSPYrYValLYaASmMeTLeThrGlyIAlaAGImeTAP	100
Db	11775	GACTGCCAGGACACGGCTGACTAGCTTAAAAACATGATCACTGGTGTCTCAATAGGAC	11834
QY	101	GIYAlaIleLeuValAlaISerAlaAlaAPGlyPProMeCProGlnThrArgIuHISile	120
Db	11895	GGCGGTACTTAGTAGATATCTGGCTGACGGTCCAAATGCCAAATCCTGTGAACACTT	11894
QY	121	LeuLeuSerArgIuValGIYValProHISileValAlaPheLeuSAnLYSGInAPMeC	140
Db	11895	CTTTATTCACGTAACTGGTGGTACCAGCACTTGTAGTATTCTTAAACAAAGTTGACATG	11954
QY	141	ValIleAPeRgInGInuLeuLeuGInuLeuValGIuMeCGLValArgGInuLeuLeuSerAla	160
Db	11955	GTTACACGATGAAGAATTTATTAAGATTATGTAAGAAATGGAAATTTGGTATTTATTAACGGA	12014
QY	161	TyrGluPheProGlyYAaPeRThrProIleValAlaGlySerAlaLeuArgAlaLeuGIn	180
Db	12015	TRTAGCTTCCAGGTGACGATGATCTGCTGATTCCTGGTTCACACTTAAAGCTTTAGAA	12074
QY	181	GIuAlaIuValaGIYaaenValGIYIuTrpGlyIuLYuValLeuLYuLeuMetAlaGlu	200
Db	12075	-----GGCGATGCTCAATACGAAGAAATAATCTTAGAATTAATGAAAGCT	12119
QY	201	ValAlaPalaTyrIleProThrProGluArgAPeRThrGluSthrPheLeuMeCProVal	220
Db	12120	GTAATATCTTACATTCACATCCACCCGAACCGATATTCGACAAACATTCATGATGCCACATT	12179
QY	221	GIuAPeRValPheSerIleAlaGlyARXGlyYThrValValThrGlyArgIleGluARXGly	240
Db	12180	GAGACGTAATTTCTCAATCACTGGTCGTCGATCTGTTGCTACAGCCGCTGTGAACGTGGT	12239
QY	241	ValValIuYsValGIYaaPGLuValGIuIleValAlaGlyIleArgProThrGlnYsThr	260
Db	12240	CAATTCAAAGTTGGTGAAGAAAGTTGAATATCATCGCTTACATGACATCTATAAACACT	12299
QY	261	ValThrGlyValAGluMeCPheArgLYuGInuLeuGInuLYuSGLYuAlaGlyaaPaaenVal	280
Db	12300	GTTACAGGTGTGAATGTTCCGTAAATTTATTAGACATCCCTGAACCTGGTGCACACATT	12359
QY	281	GIYAlaIleuLeuARXGlyYThrLYuLYuGInuGInuValAGluArgGlyMeTValLeuCYsLYs	300
Db	12360	GGTCATTTATTCAGTGGTGTGCTCGTGAAGACGTCACAGCTGTCAGATTAATGCTGCT	12419
QY	301	ProGlySerIleThrProHISLYuLYaPheGluGlyIuIleTyrValLeuSerLYSGIn	320
Db	12420	CCTGGTTCAATTACACCAACATACGANTCAAGACAGAACTATACGATTTATCAAAAGAC	12479
QY	321	GIuGlyGlyArgHISThrProPhePheThrAsnTyrArgProGlnPheTyrValAlaArg	340
Db	12480	GAAAGTGAACGTCAACATCTCTTCTCAAAACATATGTCACAAATTTATTTCCGTACT	12539
QY	341	ThrAPeRValThrGlySerIleThrLeuProGluGlyValAGluMeTValMeTProGlyAP	360
Db	12540	ACTACGTAACGTGGTGTGTTCACTTACGAAAGGATCTGAATATGCTGTGAT	12599
QY	361	AsnValLYuIleThrValGIuLeuIleSerProValAlaLeuGInuLeuGlyThrLYuSPh	380
Db	12600	AAAGTGAAGACAGATGAAATTAATGCTCCCAATCCGATTAAGAACGCTACTGCTTC	12659
QY	381	AlaIleArgGluGlyArgThrValAGlyValaGlyValaLYuSerAnIleIleGlu	399
Db	12660	TCATATCCGTGAAGGTGACGATAGTGAATGACGCGTTTACTGAATATCATTTAAA	12716

```

FILE REFERENCE: BW-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
CURRENT FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 28
LENGTH: 49617
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: incyte template ID No. 6632636 28
PUBLICATION INFORMATION:
US-09-596-002-28

Alignment Scores:
Pred. No.: 3,47e-167 Length: 49617
Score: 1570.50 Matches: 298
Percent Similarity: 86.68% Conservative: 47
Best Local Similarity: 74.87% Mismatches: 48
Query Match: 76.87% Indels: 5
DB: 4 Gaps: 2

US-10-039-183A-4 (1-399) x US-09-596-002-28 (1-49617)
QY 1 MetAlaYsGluYsPheAnaYrThLyProHiEValAsnIleGlyThrIleGlyHis 20
DB 33195 ATGGCAAAAGCGCAAGTTTGAACCGGTAAACACACAGTAATGTTGGTACACGGGTAC 3325
QY 21 ValAspHisGlyYrThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuYsGly 40
DB 33255 GTTACACCTGGTAAACACCTTACACTGCTGATTCGACACCGTTCCTGCTAAGCACAC 3331
QY 41 LeuAlaGluMetYsAspTYrAspAsnIleAspAsnAlaProGluGluYsGluYrGly 60
DB 33315 GGTGGGTGAAGCAAAAGATACGCTGCTATGACTCAGCACTTAAGAAAAGCAGCGGT 3337
QY 61 IleThrIleAlaThrSerHisIleGluTYrGluThrGluAsnArgHisTYrAlaHisVal 80
DB 33375 ATCAACATCAACACCTCTCATTTAGAGTAAGACACGTCGACCTGTCATACGCACTGTA 3343
QY 81 AspCysProGluHisAlaAspTYrValYsAsnMetIleThrGluAlaAlaGluMetAsp 100
DB 33435 GACTGCCCAAGGTCAAGCTGACTATGTTAAATAACATGATCACAGTCCGCAAGATGGAT 3349
QY 101 GluAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
DB 33495 GGTCTATCTCTGGTGTGTTCTGCAACTGAAGTGTCCTATGCGCAAACTCGTAGCATATC 3355
QY 121 LeuLeuSerArgGluValAlaGluValProHisIleValValPheLeuAsnLYsGluAspMet 140
DB 33555 CTACTATCTCGTCAAGTGGTGGTGTGACATACATATGCTATTCATGAAACAAGTCCGATATG 3361
QY 141 ValAspAspGluGluLeuLeuGluLeuValAlaGluMetGluValAlaArgGluLeuLeuSerAla 160
DB 33615 GTTATGATGTGAAGAGCTACTAGATTTGTTGAATGGAATGGAATTCGTGAATCTTATATCGAC 3367
QY 161 TyrGluPheProGluYsAspArgThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 33675 TATGATATTTCCCTGGTGTGTATGATCCCAATCATCAAAAGTTTCAGACACTAGAAAGCATTTGAAT 3373
QY 181 GluAlaYsAlaGluAsnValAlaGluTYrGluYsGluYsLYsValLeuLYsLeuMetAla 199
DB 33735 -----GGTTTGTATGTTGAATATATGAGGAGCGTCCGACGTTCTTAAGATCTGTAGAC 3378
QY 200 GluValAspAlaTYrIleProThrProGluYsAspTYrGluYsThrPheLeuMetPro 219
DB 33783 ACACCTAGACAGCTATATCCCAAGCGCTGAGCGATATGATATGATCATTTCTTGATGGCA 3384
QY 220 ValGluAspValPheSerIleAlaGlyYsGlyThrValValThrGluYrGluYsGly 239

```

Db	33843	ATTGAAGATGCTTCTCGACTCTCAGGCTCGGTGACACTGCTGACTGCTGCTGTTGAATCA	33902
Oy	240	GIyValVallyVsValGIyAspGluValGluIleValGIyLLeArgProThrGlnLysThr	259
Db	33903	GGTATTATTAAAGTTGGTGTGATGAAATTTGAATATCATCGTATCAAAACCACTGGCTAAACCC	33962
Oy	260	ThrValThrGluValGIuMetPheArgLysGluLeuGlnLysGluGluValAspAsn	279
Db	33963	ACCTGTACTGGTGTGGAAATGTTCCGTTAAACTGTGAACCAAGAGTGTGCAAGGTGAAC	34022
Oy	280	ValGIyValLeuLeuArgLysThrLysLysGluGluValGluValArgGlyMetValLeuCys	299
Db	34023	TGTGTATCTCTGTCGTGGTACTAAACCGGAGGAAGTTCAACGCGGTCAAGTACTTGA	34082
Oy	300	LysProGlySerLLeThrProHisLysLysPheGluGluGluLysLysLeuValLeuSerLys	319
Db	34083	AAACCAAGGTTTAATACACCCCAACATCAATGATTTGATCTGAAGCTTATGTACTGTAATA	34142
Oy	320	GluGluGlyGIyArgHisThrProPhePheThrAsnLysArgProGlnPheTyValArg	339
Db	34143	GAAAGAGGTGGTGTCTCACACCCACTTCTTAATAGGCTATGCGCCACAGTTCTACTCCGT	34202
Oy	340	ThrThrAspValThrGlySerLLeThrLeuProGluGluValGluMetValMetProGly	359
Db	34203	ACCAAGATGTGATGGTGGCCATCACTCAACAAGAAAGTACCAAAATGGTTAAGCTGTGT	34262
Oy	360	AspAsnValLysLLeThrValGluLeuLLeSerProValAlaLeuGluLeuGlyThrLys	379
Db	34263	GACAAATGTGAGATGAGAGTGTGAAGCTTATCCACCCATGCCATGCCATGATTAAGGTCTACGC	34322
Oy	380	PheAlaIleArgGluGluGlyGIyArgThrValGIyValAspGlyValValSerAsnLe	397
Db	34323	TTTCTCATCCCTGAAGCGGTGTCGATCCGTAGGTCTGGTGTCTATGTTTGTATGTTT	34376

```

RESULT 13
US-09-492-709A-86
; Sequence 86, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zykind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 86
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: E. Coli
US-09-492-709A-86

Alignment Scores:
Pred. No.:          5.67e-168
Score:              1555.50
Percent Similarity: 85.68%
Best Local Similarity: 73.37%
Query Match:        76.14%

DB:                  4
Gaps:                1
Length:              1185
Matches:             292
Conservative:        49
Mismatch:             52
Indels:              5
Gaps:                1

```

Oy 1 Metalalysglulyspheasnargthrlyspromhisvalasnilleglythrileglyhis 20
 :::::||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GTGTCTAAGAAATTTGACCGTACCAACCACGGTCGTTAACGTTGGTACTAATCGGCCAC 60

QY	21	ValAAPHsiGLyLysThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuYsgLy	40
Db	61	GTGAGCAACGGTAAACTACTCTGACCGGTGAAATACACACGGTACTGGCTAAACCTTAC	120
QY	41	LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluYsgLyArgLy	60
Db	121	GGCGGTGCTGCTCGTGCATCTGCACAGATCGATTAACGGCCGGAAAGAAAGCTCGTGGT	180
QY	61	IlleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal	80
Db	181	ATCACCATCAACACTCTTCACCTTGAATTCGACACCCCGACCCGCTCATTACGACACGTA	240
QY	81	AspCysProGluHisAlaAspTyrValLysAsnMetIleThrGlyAlaIleGluMetAsp	100
Db	241	GACTGCCCGGGACCGCCGACCTATGTTAAAAACATATACCGGTCTGCTCAGATGAC	300
QY	101	GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle	120
Db	301	GGCGGCATCTCTGTGTGTGCTGCTCGACTGACGGCCGATCCCGACACTGCTGAGCATC	360
QY	121	LeuLeuSerArgGluValGlyValProHisIleValValPheLeuAsnLysGluAsnMet	140
Db	361	CTGCTGGGTCTGCAGATAGGCGTTCCTGCATCATCTGTCTCTTGAACAAATCGACATG	420
QY	141	ValAspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla	160
Db	421	GTTATGACGAAGAGACTGCTGGAACTGGTTAAATGGAAGTTCTGTGAATCTGTCTCAG	480
QY	161	TyrGluPheProGluYAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu	180
Db	481	TACGACTTCCCGGGGACGACACTCGAATCGTTCGGTTCGTGCTTGAAAGGGCTGGAA	540
QY	181	GluAlaLysAlaGlyYAsnValGlyLysThrGlyGluLysValLeuLysLeuMetAlaGlu	200
Db	541	-----GGCGACCCAGATGGGAAGCAAAATCTGTGAATCTGCTGGCTTC	585
QY	201	ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal	220
Db	586	CTGGATTCCTTAATTCCTCGGACACAGAGCTGGATTTGAACAGCCGTTCTCTGCTCGCATC	645
QY	221	GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly	240
Db	646	GAAAGCGAATTCCTCAATCTCCGATGGGTACCGTTTACCAGGTGTGTGTAAGACCGGT	705
QY	241	ValValLysValGlyYAspGluValGluIleValGlyIleArgProThrGluLysThr	260
Db	706	ATCATCAAAAGTTGGTGAAGATTTGAATCGTTGGTATCAAAAGACTACGAAGACTTACC	765
QY	261	ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal	280
Db	766	TGTACTGGCGTGAATGTTCCGCAAACTGTGAGACGAAGGCCGCTCGTGTGAACGTA	825
QY	281	GlyValIleLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys	300
Db	826	GGTGTTCCTGCTCGGTGATCAAACTGAAGAAATGCAAGCTGTCTCAAGTACTGGCTAAG	885
QY	301	ProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGlu	320
Db	886	CCGGGACACCATCAACCCGACACCAAAAGTTGCAATCGAATGCAATGTAATCTGTCTCAAAAGAT	945
QY	321	GluGlyValArgHisIleThrProPhePheThrAsnTyrArgProGluPheTyrValArgThr	340
Db	946	GAAAGCGGGCGGCATACCTCGTTCTTCAAAAGGCTACCGTCCGCAAGTTTCACTTCCGTACT	1000
QY	341	ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp	360
Db	1006	ACTGACGTGACTGATACCATCTGAATCGCCGGAAGGGGTGAGATGGTAAATGCGCGGCGAC	1066
QY	361	AsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLysPhe	380
Db	1066	AACATCAAAATGGTTGTAACCTGTGATCCACCGCATCGCATGGACACAGCGTTCGCTTC	1122


```

QY 381 AlaileArgLuglyYArgThrValGlyAlaGlyValIleSerAsnIleIle 398
DB 1126 GCAATCCGGAAGCGCGCGTACCGTGGCGGCGGCGTTGTCGTAAGTCTG 1179

RESULT 14
US-09-489-039A-3648
; Sequence 3648, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3648
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3648

Alignment Scores:
Pred. No.: 6,186-168 Length: 1254
Score: 1555.50 Matches: 291
Percent Similarity: 85.18% Conservative: 48
Best Local Similarity: 73.12% Mismatches: 54
Query Match: 76.14% Indels: 5
DB: 4 Gaps: 1

US-10-039-183a-4 (1-399) x US-09-489-039A-3648 (1-1254)
QY 1 MetAlaLySGuLySPheAsnArgThrLySProHisValAsnIleGlyThrIleGlyHis 20
DB 70 GGTCTAAAGAAAATTGTAACGTACAAACCGCAGCTCAAGTGTGATCTGCGCCAC 129
QY 21 ValAspHisGlyLyThrThrLeuSerAlaAlaIleSerAlaValIleuSerLeuLySgly 40
DB 130 GTTGACATCGTGAACATCACTGCTGCTGCATCACTACACCGTCTGCTAAACCTAC 189
QY 41 LeuAlaGluMetLeuAspTyraAspAsnIleAspAsnAlaProGluGluLeuArgLy 60
DB 190 GGTGTCCTCCGTCGGCATTGACACGATGATACGCCGCCGAAAGAAACCTGCTGT 249
QY 61 IleThrIleAlaThrSerHisIleGlyLyThrGluThrGluAsnArgHisIleVal 80
DB 250 ATCACCATCAACACCTCTCAAGTGAATGACACCCGCACTGCGCACTAGCGCACGTA 309
QY 81 AspCySProGlyHisAlaAspTyraValyAsnMetIleThrGlyAlaAlaGluMetLasp 100
DB 310 GACTGCCCGGGCCAGCGCATATGTTAAACATGATACCGCGTGGCGCAGATGAGC 369
QY 101 GlyAlaIleLeuValIleSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
DB 370 GCGCGATCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 429
QY 121 LeuLeuSerArgGluValGlyValProHisIleValIlePheLeuAsnLySglnAspMet 140
DB 430 CTGCTGGGTCGTCAAGTAGGCGTTCGTAATCATCATCGTTCCTGAACAAACGCGATATG 489
QY 141 ValAspAspGlnGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
DB 490 GTTGATGACGAAGACTGCTGACACTGCTGAGATGAGAGTTCGTAACCTGCTCTCAG 549
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaGluGlu 180
DB 550 TACGATTTCCCGGGGAGACACCCGATCGTTCGTGCTGCTGCTGCTGCTGCTGCTGCTG 609
QY 181 GluAlaLyAlaGlyAsnValGlyIleThrGlyGlyGlyValLeuLeuSerLeuMetAlaGlu 200
DB 610 -----GCGGACGCAAGTGTGGAAGGAAATCATGTAATGCTGCTGCTGCTGCTGCTG 654

```

```

QY 201 ValAspAlaTyrlleProThrProGluArgAspThrGluLyThrPheLeuMetProVal 220
DB 655 CTGATACCTATATCCCGAACACAGAGCGTGATGACACACCGTTCCTGCTGCGGATC 714
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
DB 715 GAAGACGATATTCATCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 774
QY 241 ValValIleValIleAspGluValGluIleValGlyIleArgProThrGlnLyThrThr 260
DB 775 ATCATCAAGAGTGAAGGAAGTGAATCTGTTGATCAAGAAACCGCGCAACACACC 834
QY 261 ValThrGlyValGluMetPheArgPheGluLeuGluLySglnValAlaGlyAspAsnVal 280
DB 835 TGTACTGCGGTGAATGTTCCGCAACTGCTGCAAGAGCCGCTGCTGCTGCTGCTGCTG 894
QY 281 GlyValIleLeuArgGlyThrLyLeLySglnGluValGluArgGlyMetValLeuCySPlys 300
DB 895 GGTGTTCTGCTGCGGTGATCAACGTGAAGAAATCGAACGTGCTGCTGCTGCTGCTGCTG 954
QY 301 ProGlySerIleThrProHisLeuLySglnGlyGlyIleThrValLeuSerLySgln 320
DB 955 CCGGCGACATCAACCCCGCACACCAAGTTCGATCTGAAGTATCATCTGCTGCAAAAGAC 1014
QY 321 GluGlyLyArgHisIleThrPhePheThrAsnTyraSPProGlnPheTyraValArgThr 340
DB 1015 GAAGCGCGCGTCACTACCTGTTCAAGGCTACCGTCCGCACTTCACTTCCGTAAT 1074
QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
DB 1075 ACTGACGTGACTGCGACATCGAATCGCCGAGCGTGAAGATGATGCTGCTGCGGCGAC 1134
QY 361 AsnValLySglnThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLySphe 380
DB 1135 AACATCAAAATGTTGTTTACCTGATCCACCGAGTACGAGACGACGCTGCTGCTTC 1194
QY 381 AlaileArgLuglyYArgThrValGlyValIleSerAsnIleIle 398
DB 1195 GCAATCCGGAAGCGCGCGTACCGTGGCGGCGGCGTTGTCGTAAGTCTG 1248

RESULT 15
US-09-489-039A-3750/c
; Sequence 3750, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3750
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3750

Alignment Scores:
Pred. No.: 6,236-168 Length: 1260
Score: 1555.50 Matches: 291
Percent Similarity: 85.18% Conservative: 48
Best Local Similarity: 73.12% Mismatches: 54
Query Match: 76.14% Indels: 5
DB: 4 Gaps: 1

US-10-039-183a-4 (1-399) x US-09-489-039A-3750 (1-1260)
QY 1 MetAlaLySGuLySPheAsnArgThrLySProHisValAsnIleGlyThrIleGlyHis 20
DB 1126 GCAATCCGGAAGCGCGCGTACCGTGGCGGCGGCGTTGTCGTAAGTCTG 1179

```

Db 1200 GTGTCTAAAGAAAAATTGAAAGTACAAACCGCAGCTTCACGTTGGTACTATCGGCAC 1141
Qy 21 ValAspHisGlyLeuThrThrLeuSerAlaIleSerAlaValLeuSerLeuIle 40
Db 1140 GTTGACCATGTGTAACCTCTGACTGCTGCCATCACTACCGTTCTGGCTAAACCTTAC 1081
Qy 41 LeuAlaGluMetLeuAspTyrAspAsnIleAspAsnAlaProGluGluIleGly 60
Db 1080 GGTGGTTCGGCTCGCGCATTCGACAGATCGATTAACGGCCGGAAGAAAGAACTCGTGGT 1021
Qy 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
Db 1020 ATCACCATCAACACCTCTCACTGTAATATGACCCGCACTCGCCACTACCGCACCGTA 961
Qy 81 AspCysProGluHisAlaAspTyrValIleAspAsnMetIleThrGluAlaAlaGluMetAsp 100
Db 960 GACTGCCCGGCGCACCGCATATGTTTAAACATGATACCGGTGCTGCGCAGATGAC 901
Qy 101 GluAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
Db 900 GCGCGGATCTGTGCTGTGCTCGACTGACCGGCCGATGCCGACACTCGTAGCACATC 841
Qy 121 LeuLeuSerArgGluValGlyValProHisIleValPheLeuAsnIleGlnAspMet 140
Db 840 CTGCTGGGTCCGTCAAGTAGGCGCTTCCGATCATCATCGTCTGACCAATCGCATATG 781
Qy 141 ValAspAspGlnGluLeuGluIleValGluMetGluValArgGluLeuLeuSerAla 160
Db 780 GTTGATGACGAAGACTGCTGGAATGTTGATGATGAAATTCGTGAATCTGCTCTCAG 721
Qy 161 TyrGluPheProGluAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 720 TACGATTTCCGGGGGACGACACCCGATCGTTCGTGTTCTGCTCTGAAGCGCTGGAA 661
Qy 181 GluAlaIleValGlyAsnValGlyIleTyrGlyIleValLeuIleLeuMetAlaGlu 200
Db 660 -----GGCGACGACGAGTAGGAAGCGAAATCATCAAACTGGCTGCGCAC 616
Qy 201 ValAspAlaTyrIleProThrProGluArgAspThrGluIleThrPheLeuMetProVal 220
Db 615 CTGGATACCTATATCCCGAACCAAGCGTGCATTGACAAAGCGTTCCTGCTGCCGATC 556
Qy 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
Db 555 GAAAGCGATATCTCATCTCCGCTGTGTGTAACGTTGTTACCGCTGCTGTAAGCGCGGT 496
Qy 241 ValValIleValGlyAspGluValGluIleValGlyIleArgProThrGlnIleThrThr 260
Db 495 ATCATCAAAAGTAGGAAGAGTTGAAATCGTTGGTATCAAAAGAAACCGCGAAACCAAC 436
Qy 261 ValThrGluValGluMetPheArgIleGluGluIleGluGluIleGluIleValAspAsnVal 280
Db 435 TGTACTGGCGTTGAATGTTCCGCAAACTGCTGACGAAGCGCTGTGTGTGAGAACGTA 376
Qy 281 GluValLeuLeuArgGlyThrIleIleGluGluIleValGluArgGlyMetValLeuCysIle 300
Db 375 GGTGTTCTGCTGCGGTGATCAACGTGAAGAAATCGAACGTGCTCAGTACTGCTAAG 316
Qy 301 ProGlySerIleThrProHisIleIlePheGluGluIleTyrValLeuSerIleGlu 320
Db 315 CCGGCGACCATCAACCCGACACCAAGTTCCGAATCTGAAGTGTATCATCTCTGCCAAAGAC 256
Qy 321 GluGlyGlyArgHisIleThrProPhePheThrAsnTyrArgProGluIlePheTyrValArgThr 340
Db 255 GAAAGCGGCGCGTCACTCCGTTCTTCAAGGCTACCGTCCGACGTTCTACTTCCGTA 196
Qy 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 195 ACTGACGTGACTGGACCATCGAATCGCCGGAAGCGTAGAGATGTTAATGCCGGCGAC 136
Qy 361 AsnValIleIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrIleAsp 380
Db 135 AACATCAAAATGTTTACCTGATCAACCGGATCGCATGACGACGCGTCTGCGTTTC 76

Qy 381 AlaIleArgGluGlyIleArgThrValGlyValAlaGlyValIleSerAsnIleIle 398
Db 75 GCATTCGTTGAAGCGGCGGTATCCGTTGCGCGCGGCGGCTGTGTAAGCTAAAGTTCTG 22

Search completed: April 24, 2005, 12:23:55
Job time : 2662.22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2005, 05:52:31 ; Search time 777.993 Seconds

(without alignments)
3113.588 Million cell updates/sec

Title: US-10-039-183a-4

Perfect score: 2043

Sequence: 1 MAKEKFRNTRKPHVNIPTIGH.....FAIREGRTRVAGVVSNIIE 399

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10039183/runat_22042005_122426_20167/app_query.fasta_1.1038
-DB=Published Applications NA -QMT=fastcap -SUFFIX=irpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pcpt -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcpt -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10039183@cgn2_1.1.894@runat_22042005_122426_20167
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2043	100.0	1200	9	US-09-815-242-7318
2	2043	100.0	1200	17	US-10-282-122A-22730
3	2043	100.0	1448	8	US-08-831-310-3
4	2043	100.0	1448	15	US-10-039-183A-3
5	2035	99.6	2825	19	US-10-662-126-3
6	1772	86.7	1200	17	US-10-282-122A-18144
7	1608	78.7	1185	9	US-09-815-242-3832
8	1608	78.7	1185	17	US-10-282-122A-6275
9	1608	78.7	1185	9	US-09-815-242-6385
10	1606.5	78.6	1185	17	US-10-282-122A-29444
11	1606.5	78.6	1185	17	US-10-282-122A-29456
12	1598.5	78.2	1182	17	US-10-282-122A-29230
13	1597.5	78.2	1188	17	US-10-282-122A-11657
14	1592.5	77.9	1188	17	US-10-282-122A-14300
15	1592	77.9	684707	17	US-10-398-221-9
16	1592	77.9	3011208	17	US-10-398-221-2058
17	1586	77.6	1188	17	US-10-282-122A-24743
18	1585.5	77.6	1186	17	US-10-282-122A-29207
19	1581.5	77.4	1188	17	US-10-282-122A-14613
20	1581.5	77.4	1188	17	US-10-282-122A-14772
21	1579.5	77.3	1185	9	US-09-815-242-6562
22	1579.5	77.3	1185	9	US-09-815-242-6572
23	1579.5	77.3	1185	17	US-10-282-122A-22041
24	1579.5	77.3	1185	17	US-10-282-122A-22060
25	1579.5	77.3	1830121	17	US-10-329-670-1
26	1579.5	77.3	1830121	17	US-10-329-670-1
27	1579.5	77.3	1830121	18	US-10-158-865-1
28	1579.5	77.3	1830121	18	US-10-158-865-1
29	1578.5	77.3	1182	17	US-10-282-122A-35489
30	1576.5	77.2	1188	17	US-10-282-122A-12928
31	1572.5	77.0	1182	17	US-10-282-122A-34285
32	1572.5	77.0	1185	17	US-10-282-122A-31070
33	1572	76.9	4315	9	US-09-070-927A-1
34	1571.5	76.9	1182	9	US-09-815-242-4519
35	1571.5	76.9	1182	19	US-10-470-0488-115
36	1571.5	76.9	1185	9	US-09-815-242-8344
37	1571.5	76.9	1185	17	US-10-282-122A-7572
38	1571.5	76.9	1185	19	US-10-857-625-240
39	1571.5	76.9	4555	19	US-10-857-625-19
40	1571.5	76.9	15598	8	US-08-781-986A-82
41	1571.5	76.9	15598	17	US-10-329-624-82
42	1571.5	76.9	15901	19	US-10-857-625-16
43	1570.5	76.9	1182	17	US-10-282-122A-32228
44	1570.5	76.9	1188	17	US-10-282-122A-26722
45	1570.5	76.9	49617	17	US-10-672-787-28

ALIGNMENTS

RESULT 1
US-09-815-242-7318
Sequence 7318, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Twilick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA, 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7318
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (1200)
US-09-815-242-7318

```

```

Alignment Scores:
Pred. No.: 6,78e-187 Length: 1200
Score: 2043.00 Matches: 399
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

```

US-10-039-183a-4 (1-399) x US-09-815-242-7318 (1-1200)

```

QY 1 MetAlaLysGluYpHeAsnArGThrIysProHISValAsnIleGlyThrIleGlyHis 20
DB 1 ATGGCAAAAGAAAGATTAAACAGAACTAAGCCGATGTTAAATATGGAAACCATGGGCAT 60
QY 21 ValAspHisGlyLysThrThreusSerAlaIleSerAlaValIleuSerLeuLysGly 40
DB 61 GTAGCCATGCTGTAACCACTTGTAGTGCAGCGATTTCAAGCGTCTTTCTTGAAGGT 120
QY 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluYpGluArgGly 60
DB 121 CTTCGAGAAATGAAGACTATGATATATGTAACGCCCTGAGAAAGAAAGAGGG 180
QY 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
DB 181 ATCACTATCGCTACTTCTCACTTAATATGAGACTGAAACAGACACTATGCCATGTG 240
QY 81 AspCysProGlyHisAlaAspTyrValIysAsnMetIleThrGlyAlaAlaGluMetAsp 100
DB 241 GATTGCCAGACAGACCTGACTATGTAAACATGATACCGCGTCCGCGCAATGAGAC 300
QY 101 GlyAlaIleLeuValIleSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
DB 301 GGAGGAGATTGGTGTCTTCGACACTGAGCCCTTAAGCCCTCAAACTAGGAGCATATC 360
QY 121 LeuLeuSerArgGlnValGlyValProHisIleValIleValPheLeuAsnLysGlnAspMet 140
DB 361 TTATTGCTCGCAAGTAGGCGCTGCTCAATCGTGTGTTCTTTAAACAAACAAACATG 420
QY 141 ValAspAspGlnGluLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
DB 421 GTAGTAGCAAGAAATTTGTAGAACTTTGAAATGAGAAATGCGGAATTTGTAGCGCG 480
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 481 TATGATATTTCTCGGAGTAGACCTCTATCTAGAGCGGTTGCTTTAAGACTTTAGAA 540
QY 181 GluAlaLysAlaGlyAsnValGlyGluTyrGlyGluYpValLeuLysLeuMetAlaGlu 200
DB 541 GAAGCAAAAGCGTGAATGTGGTGAATGGGTGAAAAAGTGTCTTAACTTATGCTGAA 600

```

```

QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluYpThrPheLeuMetProVal 220
DB 601 GTGATGCTTATATCCCTACTCCAGAAAGACACTGAAAAAATCTTCTTATGCGCGTT 660
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValIThrGlyArgIleGluArgGly 240
DB 661 GAAGATGTGTTCTCTATTTGGCGGTAGAGGAGCTGTGTGTAAGGTGAAGGTGAAGAGAG 720
QY 241 ValIleLysValIleGlyAspGluValIleValIleGlyIleAcPProThrGlnLysThrThr 260
DB 721 GTGCTGAAGTAGAGCGATGAAGTGAATCTTGTGTATCAACCTACACAAAAACAGACT 780
QY 261 ValThrGlyValIleGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal 280
DB 781 GTAACCGGTGTAGAAATGTTTAGAAAGAGTTGAAAAAGGTGAAGCCGCGCATATGTG 840
QY 281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluAlaGlyLysMetValLeuCysLys 300
DB 841 GCGGTCTTTTGAAGAGCACTAAAAAAGAGAGTGAACCGGTATGTGTTCTATGCAAA 900
QY 301 ProGlySerIleThrProHisLysLysAspPheGluGluGluIleTyrValIleuSerLysGlu 320
DB 901 CCGGTTCTATCACTCCGCAAGAAATTTAGGAGAAATTTATGTTCTTTCTTAAGAA 960
QY 321 GluGlyArgGlyHisThrProPhePheThrAsnTyrArgProGluPheTyrValArgThr 340
DB 961 GAAGCGCGGAGACACACTCCATTTCTTCAACATTACCGCCGCAATTTCTATGTGGCACA 1020
QY 341 ThrAspValIThrGlySerIleThrLeuProGluGluValIleGluMetValMetProGlyAsp 360
DB 1021 ACTGATGTGACTGCTCTATCACTTCTTCAAGCGGTGAAGATGTATATCTGTGGCAT 1080
QY 361 AsnValLysIleThrValIleGluLeuIleSerProValAlaLeuGluLeuGlyThrLysPhe 380
DB 1081 AATGTGAATAATCACTGTAGATTGATTTAGCCCTGTGTGGTTAGACTGGAACTAAATTT 1140
QY 381 AlaIleArgGluGlyGlyArgThrValIleValIleGlyValIleSerAsnIleGlu 399
DB 1141 GCGATTCGTGAAGCGGTGAGACCGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGAA 1197

```

RESULT 2

US-10-282-122A-22730

; Sequence 22730, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Twick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FID REFERENCE: ELITRA.03A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

US-10-039-183A-4 (1-399) x US-10-282-122A-22730 (1-1200

```

Db      661 GAAAGTGTCTCTTAATGCGGGTAGAGGAAGACTGTGTACAGGTAGCATTTGAAAAGGC   720
Oy      241 ValVallyValIGlyAspGluValGlnIleValIGlyIleArgProThrGlnYSerThr    260
Db      721 GTGTGAAAGAAGAGCGAATGAAGTGAATACTGTGTATCAGCACTAACAAAAACGACT     780
Oy      261 ValThrGlyValIGlMetPheArgHysGluLeuGluLysGlyGlnIleArgIleAspAsnVal    280
Db      781 GTAACCGGTGTCGAAATGTTTTTGNNAGAGTTGGAAAAAGGTGMACC CGCGCTAATTGTG     840
Oy      281 GlyValIleLeuSerArgLYThrLybLySGluGluValIGluArgGlyMetValLeuCyLys    300
Db      841 GGCGTGCTTTTGAGAGGAAGTAATAAAAAAGAAAGTGAAGCGCGTATGGTTCTATCAAA     900
Oy      301 ProGlySerIleThrProHisLybAlbYbPheGluGlyGlnIleTyValLeuSerLybGlu    320
Db      901 CCAGGTTCTATCACCTCCACACAAGAAATTTGAGGGAATTAATCTCTTCTAAAGAA     960
Oy      321 GluGlyGlyValArgHleThrProPheThrAnTYrArgProGlnPheTYrValArgThr     340
Db      961 GAAGCGGGAGACACACTCCATCTTCACCAATTACCGCCGCCAATTCATGTGCGACA    1020
Oy      341 ThrAspValThrGlySerIleThrLeuProGluGlyValIGlMetValMetProGlyAsp     360
Db      1021 ACTAGTGTGACTGGCTCTATACCTCTTCTGGAAGCGTGAAGATGTTATGCTGCGAT    1080
Oy      361 AsnValLybIleThrValIGluLeuIleSerProValAlaLeuGluLeuGlyThrLybPhe     380
Db      1081 AATGGAATAATCACCTGTAGAGTTGATTAAGCCCTGTGCGTTAGAGTGGACAAATAATTT    1140
Oy      381 AlaIleArgGluGlyValArgThrValGlyValIGlyValIleSerIleIleGlu     399
Db      1141 GCGATTCGTGAAGGCGGTGAGACCGTGTGCTGTGTGTGAGCAATATTATTGAA    1197

RESULT 3
US-08-831-310-3
; Sequence 3, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleantchous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPo 1360 and
; TITLE OF INVENTION: GHPo 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Ribling LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831.310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

MOLECULE TYPE: Genomic DNA
 FEATURE:
 NAME/KEY: Coding Sequence
 LOCATION: 118...1314
 OTHER INFORMATION:
 US-08-831-310-3

Alignment Scores:

Pred. No.:	8,586-187	Length:	1448
Score:	2043.00	Matches:	399
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-10-039-183A-4 (1-399) x US-08-831-310-3 (1-1448)

```

QY      1 MetAlaLySGluYbPheAsnArgThrLySProHISValAsnIleGIYThrIleGIYHis 20
DB      118 ATGGCAAAAGAAAAGTTTAAAGAACTAAGCCGATGTTAATATTGAAACCATGGGCAT 177
QY      21 ValAspHisGIYLySThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLySGly 40
DB      178 GTAGACCATGTGTAAGACGCTTGAGTGCAGCATTTTCAGCGGTCTTCTTGTGAAGGT 237
QY      41 LeuAlaGIuMetLyAspTYrAspAsnIleAspAsnAlaProGIuGIuLySGluArgGIY 60
DB      238 CTTCAGAAATGAAGACTATGATTAATTTGTTAAACCCCTGAAAGAAAAGAAAGAGG 297
QY      61 IleThrIleAlaThrSerHisIleGIuTYrGIuThrGIuAsnArgHisTYrAlaHisVal 80
DB      298 ATCACTATCGCTACTTCTCAATGATATGAGATGAAAAAGACACATATCGCATGTG 357
QY      81 AspCysProGIuYHisAlaAspTYrValLySAsnMetIleThrGIYAlaAlaGIuMetAsp 100
DB      358 GATTGCCCAAGACACGCTGATCATGTATTAATAAACATGATCACGCGCGCAAAATGAGC 417
QY      101 GIYAlaIleLeuValValSerAlaAlaAspGIYProMetProGIuThrArgGIuHisIle 120
DB      418 GAGAGCATTTTGTTGTTCTTCTGACGTAGTGCCTTAATGCCCTCAAACTHGGAGCATATC 477
QY      121 LeuLeuSerArgGIuValGIYValProHisIleValPheLeuAsnLySGlnAspMet 140
DB      478 TTATTGTCCTCGCAAGTAGAGCGTGCCTCATCGTTGTTCTTAAACAAACAGACATG 537
QY      141 ValAspArgGIuGIuLeuGIuLeuGIuValGIuMetGIuValArgGIuLeuLeuSerAla 160
DB      538 GTAGATGACCAAGAAATGTTAGAACTTGTAAGAAATGGAAGTCCGGAATTTGTAGCCG 597
QY      161 TYrGIuPheProGIYAspAspThrProIleValAlaIleSerAlaLeuArgAlaLeuGIu 180
DB      598 TATGAATTTCTCGGCGATGACACTCTATCTGAGCGGTTGAGCTTTTAAGACTTTAGAA 657
QY      181 GIuAlaLySAlaGIYAsnValGIYGIuTYrGIYGIuLySValLeuLySLeuMetAlaGIu 200
DB      658 GAAGCAAAAGCGCTGTAATGTGGGTGAATGGGGTGAATAAGTCTTAAACTTATGCTGAA 717
QY      201 ValAspAlaTYrIleProThrProGIuArgAspThrGIuLySThrPheLeuMetProVal 220
DB      718 GTGATGCTCTATATCTCACTCCAGAAAGACACTGAAAAAACTTTCTGTATGCGGTT 777
QY      221 GluAspValPheSerIleAlaGIYArgGIYThrValIleThrGIYArgIleGIuArgGIY 240
DB      778 GAAAGTGTCTTCTTAATTCGGGTAGAGGACTGTGTTACAGGTAGATTAAGAAAGAGGC 837
QY      241 ValValLySValGIYAspGIuValGIuIleValGIYIleArgProThrGIuLySThrThr 260
DB      838 GTGGGAAGATGAGGAGATGAGTGAATCGTTGGATACAGCCCTAACCAAAAAAGCACT 897
QY      261 ValThrGIYValGIuMetPheArgLySGluLeuGIYLySGluAlaGIYAspAsnVal 280
DB      898 GTAAACCGGTGTAAGAAATGTTAGAAAGGTTGAAAAAGGTGAAGCCGCGCATATATGTG 957

```

```

QY      281 GlyValLeuLeuArgGIYThrLySLeSGluGIuValGIuArgGIYMetValLeuCysLyS 300
DB      958 GCGGTGCTTTTGAGAGGAACCTAAAGAAAGAGAGGACCGGATATGTTCTATGCAAA 1017
QY      301 ProGIYSerIleThrProHisIleLySPhelGIuGIuIleTYrValLeuSerLySGlu 320
DB      1018 CCAGGTCTATCACTCCGCCACAGAAATTTAGAGGAGAAATTTATGTCCTTCTTAAGAA 1077
QY      321 GluGIYGIYArgHisIleThrPhePheThrAsnTYrArgProGIuPheTYrValArgThr 340
DB      1078 GAAGCGGAGACACACTCTCAATTTCTTACCAATTAACCGCCGCAANTTTATGTGCGACA 1137
QY      341 ThrAspValThrGIYSerIleThrLeuProGIuGIYValGIuMetValMetProGIYAsp 360
DB      1138 ACTGATGACATGGGTCTATCACTCTTCCTGAAGCGGTGAAGATGTTATGCTTGGCAT 1197
QY      361 AsnValLySIIeThrValGIuLeuIleSerProValAlaLeuGIuLeuGIYThrLySPh 380
DB      1198 AATGGAATAACACTGTAGAGTTGATTAAGCCCTGTGGGTGAGATTGGGAACATAAATT 1257
QY      381 AlaIleArgGIuGIYGIYArgThrValGIYAlaGIYValIleSerAsnIleIleGIu 399
DB      1258 CGATTCTGTAAGCGGTGAGACCGTTGCTGTGTGTGAGCAATATTATGAA 1314

```

RESULT 4

US-10-039-183A-3

Sequence 3, Application US/10039183A

Publication No. US20030143242A1

GENERAL INFORMATION:

APPLICANT: Kleantous, Harold

APPLICANT: Liesolo, Ling

APPLICANT: Tomb, Jean-Francois

APPLICANT: Miller, Charles

APPLICANT: Al-Garawi, Amal

TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750

TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules

FILE REFERENCE: 06132/037002

CURRENT APPLICATION NUMBER: US/10/039, 183A

PRIOR FILING DATE: 2002-12-13

PRIOR APPLICATION NUMBER: US 06/831, 310

PRIOR FILING DATE: 1997-04-01

NUMBER OF SEQ ID NOS: 18

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 1448

TYPE: DNA

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: CDS

LOCATION: (118) ... (1314)

US-10-039-183A-3

Alignment Scores:

Pred. No.:	8,586-187	Length:	1448
Score:	2043.00	Matches:	399
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-10-039-183A-4 (1-399) x US-10-039-183A-3 (1-1448)

```

QY      1 MetAlaLySGluYbPheAsnArgThrLySProHISValAsnIleGIYThrIleGIYHis 20
DB      118 ATGGCAAAAGAAAAGTTTAAAGAACTAAGCCGATGTTAATATTGAAACCATGGGCAT 177
QY      21 ValAspHisGIYLySThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLySGly 40
DB      178 GTAGACCATGTGTAAGACGCTTGAGTGCAGCATTTTCAGCGGTCTTCTTGTGAAGGT 237
QY      41 LeuAlaGIuMetLyAspTYrAspAsnIleAspAsnAlaProGIuGIuLySGluArgGIY 60
DB      238 CTTCAGAAATGAAGACTATGATTAATTTGTTAAACCCCTGAAAGAAAAGAAAGAGG 297

```

```

Qy 11EthrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
Db 298 ATCACTATCGCTACTCTTCCACATTGATGAGATGAAAAACAGCACTATCGCATGTG 357
Qy 81 AspCybProGlyHisAlaAspTyrValIysAsnMetIleThrGlyAlaAlaGlnMetAsp 100
Db 358 GATTGCCAGGACACGCTGACTATGTAATAAAAAATGATCACCGGCGCGCAATGTGAC 417
Qy 101 GlyAlaIleLeuValIleSerAlaAlaAspGlyProMetProGlnThrArgGlnHisIle 120
Db 418 GGAGGATTTGGTGTGTTCTGACACTGATGGCCCTTAAGCCCAACAGGAGCATATC 477
Qy 121 LeuLeuSerArgGlnValAlaGlyProHisIleValIlePheLeuAsnIleGlnAspMet 140
Db 478 TTATTGCTCTCGTCAAGTACGAGCGCTGCTCATCGTTGTTTCTTAACAAACAGACATG 537
Qy 141 ValAspAspGlnGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
Db 538 GTAGATGACCAAGATTGTTAGAACTTGAGAAATGGAAGTCCGGAATTGTGAGCCG 597
Qy 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 598 TATGAAATTCCTGGCGATGACACTCTCATCGTACGGGCTTCAGCTTTAAGACTTTAGAA 657
Qy 181 GluAlaIysAlaGlyAsnValAlaGluTyrGlyGlyValIleLeuLeuMetAlaGlu 200
Db 658 GAAGCAAAAGCGCTGTAATGTGGTGAATGGGTGAAAAAGCTTAACTTAATGGCTAA 717
Qy 201 ValAspAlaTyrIleProThrProGluArgAspThrGlyValThrPheLeuMetProVal 220
Db 718 GTGAGTGCCTATATCCCTACTCCAGAAAGACATGAAAAAACTTTCTTGATGCCGTT 777
Qy 221 GluAspValPheSerIleAlaGlyArgGlyThrValIleThrGlyArgIleGluArgGly 240
Db 778 GAAGATGTGTTCTCTATTCGGGTAGAGGAGCTGTGGTTACAGTGAAGTGAAGAGAGC 837
Qy 241 ValIleIysValAlaGlyAspGluValGluIleValGlyIleArgProThrGlnIysThrThr 260
Db 838 GTGGTGAAGATGAGCGATGAAGTGAATTCGTGGTATCAGCTTACCAAAAAACGACT 897
Qy 261 ValThrGlyValGluMetPheArgGlyGluLeuGluGlyGlyGluAlaGlyAspAsnVal 280
Db 898 GTAACCGGTGTGAATGTTTAGAAGAGTTGGAAGAAAGTGAAGCCGCGCATATGTG 957
Qy 281 GlyValIleLeuLeuArgGlyThrIleValGluGluValGluArgGlyMetValIleuCyAlys 300
Db 958 GCGCGTCTTTGAGAGAACTAAAGAAAGAGAGAAAGCGGATGCTTATGCAAA 1017
Qy 301 ProGlySerIleThrProHisIleValIysPheGluGluIleValIleLeuSerIleGlu 320
Db 1018 CCAGGTGTATCATCTCCGACAAAGAAATTTGAGGAGAAATTTAGTCTTTCTAAAGAA 1077
Qy 321 GluGlyGlyArgHisThrProPheThrAsnTyrArgProGlnPheTyrValArgThr 340
Db 1078 GAAGCGGAGACACACTCCATCTTCAACAAATTCACCGCCGCAATTTCTATGTGCGACA 1137
Qy 341 ThrAspValIleThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 1138 ACTGATGTGACTGCTCATCATCCCTCTCGAAGCGGTAGAAATGTTATGCTGCGCAT 1197
Qy 361 AsnValIysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrIysPhe 380
Db 1198 AATGTGAATAATCATCTGATGAGTGTAGTCCCTGTGGTGAAGTGTGGGAACTAAATTT 1257
Qy 381 AlaIleArgGluGlyArgThrValAlaGlyValIleValIleSerAsnIleIleGlu 399
Db 1258 GCGATTCTGTGAAGCGGTAGAACCTTGCTGCTGTGTGACGAAATTTATTGAA 1314

```

RESULT 5
US-10-662-126-32
; Sequence 32, Application US/10662126
; Publication No. US20050063987A1

```

; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hündt, Erika
; APPLICANT: Schmidt, Karl-Heinz
; TITLE OF INVENTION: Proteins, in Particular Membrane Proteins, of
; FILE REFERENCE: CHIR-0340
; CURRENT APPLICATION NUMBER: US/10/662,126
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: 09/230,158
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: PCT/IB97/00981
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 2825
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (891)..(2090)
; OTHER INFORMATION: bp protein; 42 kd protein from Helicobacter pylori
US-10-662-126-32

Alignment Scores:
Pred. No.: 1,17e-185 Length: 2825
Score: 2035.00 Matches: 398
Percent Similarity: 99.75% Conservative: 0
Best Local Similarity: 99.75% Mismatches: 1
Query Match: 99.61% Indels: 0
DB: 19 Gaps: 0

US-10-039-183a-4 (1-399) x US-10-662-126-32 (1-2825)
Qy 1 MetAlaIysGluYsPheAsnArgThrIysProHisValIleGlyThrIleGlyHis 20
Db 891 ATGGCAAAAGAAAAGTTTAAACAGAACTAAGCGCATGTTATATGGAACCATGGCAT 950
Qy 21 ValAspHisGlyIysThrThrLeuSerAlaAlaIleSerAlaValLeuSerIleuYsGly 40
Db 951 GTAGCCCATGTGTAAACACACTTGTGACGACGATTCAGCGGCTCTTTTGAAAGGT 1010
Qy 41 LeuAlaGluMetIysAspTyrAspAsnIleAspAsnIleProGluGluYsGluArgGly 60
Db 1011 CTGAGCAATGAAAGATGATGATATATGATACGCCCTGAGAAAGAAAGAGAGG 1070
Qy 61 11EthrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
Db 1071 ATCACTATCGCTACTCTTCCACATTGATGAGATGAAAAACAGCACTATCGCATGTG 1130
Qy 81 AspCybProGlyHisAlaAspTyrValIysAsnMetIleThrGlyAlaAlaGlnMetAsp 100
Db 1131 GATTGCCAGGACACGCTGACTATGTAATAAAAAATGATCACCGGCGCGCAATGTGAC 1190
Qy 101 GlyAlaIleLeuValIleSerAlaAlaAspGlyProMetProGlnThrArgGlnHisIle 120
Db 1191 GGAGGATTTGGTGTGTTCTGACACTGATGGCCCTTAAGCCCAACAGGAGCATATC 1250
Qy 121 LeuLeuSerArgGlnValAlaGlyProHisIleValIlePheLeuAsnIleGlnAspMet 140
Db 1251 TTATTGCTCTCGTCAAGTACGAGCGCTGCTCATCGTTGTTTCTTAACAAACAGACATG 1310
Qy 141 ValAspAspGlnGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
Db 1311 GTAGATGACCAAGATTGTTAGAACTTGAGAAATGGAAGTCCGGAATTGTGAGCCG 1370
Qy 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 1371 TATGAAATTCCTGGTGAAGACACTCTCATCGTACGGGCTTCAGCTTTAAGACTTTAGAG 1430
Qy 181 GluAlaIysAlaGlyAsnValAlaGluTyrGlyGlyValIleLeuLeuMetAlaGlu 200

```

Db 1431 GAAGCAAGGCTGTATGTGGTGAATGGGGTGAAGAAAGCTTAAGCTCAGCTGAA 1490
Qy ValaaspalaTyrlleProthProgluArgspThrgluVstThrPheleuMeProval 220
Db 1491 GTGATGCTTATATCCCTACTCCAGAAAGACACTGAAGAAACCTTCTTGATGCGGTT 1550
Qy 221 GluaspValPheSerlleAlaGlyArgGlyThrValValThrglyArglleGluArgGly 240
Db 1551 GAAGATGTGTTCTCTATTGCGGGTGAAGGAGCTGCGTTACAGGTAAGATTGAAGAGGT 1610
Qy 241 ValVallyValGlyAspGluValGluIleValGlylleArgProThrglnlysthrThr 260
Db 1611 GTGGTGAAGATGAGCGATGAAGTGAAGATCGTTGTATCAGCTACCAAAAAAGACT 1670
Qy 261 ValThrglyValGluMeCPheArglysgluValGluValGluArgGlyAspVal 280
Db 1671 GTTACCGGTGTGAATGTTTGAAGAAAGCTGAAGAAAGGAGGCGCGCATATATGTG 1730
Qy 281 GlyValleuLeuArgGlyThrglysgluValGluArgGlyMeValleuCylys 300
Db 1731 GCGGTGCTTTGAGAGGAACATAAAAAAGAAAGTAAGAACGCGGTATGTTCTATGCAAA 1790
Qy 301 ProGlySerlleThrProHlsyBlyspheGluGlyIleTyrValleuSerlysglu 320
Db 1791 CCAAGTTCTATCATCCCGACAAAGAAATTGAGGAGAAATTATGCTCTTTCTAAAGAA 1850
Qy 321 GluGlyArgGlyThrPhePheThrAsnTyrArgProGluPheTyrValArgThr 340
Db 1851 GAAGCGGGAGACACATCTCATTTCTACCAATACCGGCCCAATTTCTATGTGGCAGC 1910
Qy 341 ThraSpValThrglySerlleThrleuProgluGlyValGluMeValMeProgluAsp 360
Db 1911 ACTGATGTGACTGGGCTCATCACTCTCTGAAGGGGTGAAGATGTTATGCTGCGCAT 1970
Qy 361 AsnVallySerlleThrValGluIleuIleSerProValAlaGluGluGlyThrysphe 380
Db 1971 AATGGAATAATCAGTGAAGTGAATGATGCCCTGTGCGTTAGAGTGGGAACATAATTT 2030
Qy 381 AlaIleArgGluGlyValArgThrValGlyValGlyValSerAsnIleIleGlu 399
Db 2031 GCGATTCGTGAAGCGCGTGAAGCCGTTGTGCTGTGTTGTAGCAATATTTGA 2087
RESULT 6
US-10-282-122A-18144
; Sequence 18144, Application US/10282122A
; Publication No. US20040029129A1
GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18144
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
US-10-282-122A-18144
Alignment Scores:
Pred. No.: 8.19e-161 Length: 1200
Score: 1772.00 Matches: 336
Percent Similarity: 93.23% Conservative: 36
Best Local Similarity: 84.21% Mismatches: 27
Query Match: 86.74% Indels: 0
DB: 17 Gaps: 0
US-10-039-183A-4 (1-399) x US-10-282-122A-18144 (1-1200)
Qy 1 MecAlaLySGluVspPheAsnArgThryspProHlsValAsnIleGlyThrlleGlyHis 20
Db 1 AATGCTAAAGAAATTTTACGTATAGCCACACGTAATATTTGGTACTATGTCAT 60
Qy 21 ValAspHisGlyVstThrThrleuSerAlaAlaIleSerAlaValleuSerleuGly 40
Db 61 GTTGAACATGTGAATAACTATTAAACAGCTGCTATTTCTGCTGTTCTTTCTTAAGAGAGT 120
Qy 41 LeuAlaIleuMeLysAspTyrAspAsnIleAspAsnAlaProgluGluVstGluArgGly 60
Db 121 TTAGCAGAGCTTAAAGATTATGATATATCGATATATGCTCCAGAAAGAAAGAGCTGCT 180
Qy 61 IleThrlleAlaIleThrSerHisIleGluTyrGluThrgluAsnArgHisTyrAlaHisVal 80
Db 181 ATTACTATATGCTACTCTCATATTTGAATGATAAACAAGACAAATCGCTATATGCAACGTT 240
Qy 81 AspCySProGlyHisAlaAspTyrVallyAsnMeCilleThrglyAlaAlaGluMeAsp 100
Db 241 GACTGCCAGGTCACGCGAGATTATGTTAAACATGATTACAGGTGCTGCAAAATGAGT 300
Qy 101 GlyAlaIleleuValValSerAlaAlaAspGlyPromeProGlnThrArgGluHisIle 120
Db 301 GGAAGGATCTTGTTGTTCTGCTGCAAGATGAGCCCTTAAGCCAAACTAGAGACATTT 360
Qy 121 LeuLeuSerArgGluValGlyValProHlsIleValValPheLeuAsnlyGluAspMeC 140
Db 361 CTTCCTTCTCGTCAAGTGAAGGCGTTCCATATATTTGTTTATGAATAAAGCAGATATG 420
Qy 141 ValAspAspGluGluIleuLeuGluIleuValGluMeCylValArgGluLeuLeuSerAla 160
Db 421 GTTGAATGCTGAACCTTTAGAGTATGATTAAGAAATTAAGAAATTAAGAAATTAAGCTCT 480
Qy 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaAlaArgAlaLeuGlu 180
Db 481 TATGATTTCCAGGAGGAGATACACTATATTTCTGCTTCTGCTTTAAAGCTCTTGA 540
Qy 181 GluAlaLySAlaGlyAsnValGlyGluTyrGlyGluValleuLySLeuMeCAlaGlu 200
Db 541 GAAGCTAAAGCTGAGCAAGATGTGAATGTGACAGAAATAATTAAGATCTTATGCTTCA 600
Qy 201 ValaSpAlaTyrlleProthProgluArgspThrgluVstThrPheleuMeProval 220
Db 601 GTTGAATGCTATATTTCCAACTCAACTGTAATCTGAATAAAGAACTTCTTATGCAATTT 660


```

QY      .221 giuaapvalpaseserilealaaglyarglythrvalvalthrvalargylegulargly 240
DB      GAAAGATGTTTCTCAATTCAGGTGCTGTACTGTTTACAGGTGAATGAAAAAGCT 720
QY      241 valvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalval 260
DB      721 GTTGTAAAGATGATGATCTATCGAAATCGTTGGATTAAAGATCTCAACACACACT 780
QY      261 valthrvalvalglumephearglysgluvalvalvalvalvalvalvalvalvalval 280
DB      781 GTPAACAGGTTGTGAATGTTCAAGAAAGAAATGATCAAGGCAAGCAAGATTAACGTA 840
QY      281 giualvalleuvalarglythrvalvalvalvalvalvalvalvalvalvalvalvalval 300
DB      841 GGTGTTCTTCTTCGTGTACTTAAAGAAAGAAAGATTAACGTTGATGTTCTTCTTAA 900
QY      301 Proglyserilethrprohislylsyphleglygluilethyvalleuserlysglu 320
DB      901 CCAAAATCAATTACTCCACACTGACTCGAAGCTGAAGTTTATCTTAATAATAAGAT 960
QY      321 giulglvalarghilethrprophethrasmtyrargprogluphettyrvalargthr 340
DB      961 GAAGGTGATGACATATCTCATCTTTAAACAATATAGACACAGTTTATGTAAAGAA 1020
QY      341 Thraservalthrvalserilethrleuproglyvalglumetvalmetproglyaap 360
DB      1021 ACTGATGTTACAGTTCCGATTAATTAAGTATGCTGTTGAATGTTTATGCCAGTGA 1080
QY      361 Aanvalvalilethrvalgluleuileserprovalaleuugluuileuglythrlyspe 380
DB      1081 AATGACAGAAATTAAGTATGATGCTGCTCAGTACGACTGAAGAAAGAACTCGTTT 1140
QY      381 Aalilearglygluylarghthrvalglualaglyvalvalserasmlleleu 399
DB      1141 GCTATTCGGAAGGTGTAACCTGTGTTCAAGTTGTTCTTAATAATTATTA 1197

RESULT 7
US-09-815-242-3832
/ Sequence 3832, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zykkind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3832
/ LENGTH: 1185
/ TYPE: DNA

```

```

/ ORGANISM: Enterococcus faecalis
US-09-815-242-3832

Alignment Scores:
Pred. No.: 4,9e-145 Length: 1185
Score: 1608.00 Matches: 308
Percent Similarity: 87.25% Conservative: 41
Best Local Similarity: 77.00% Mismatches: 45
Query Match: 78.71% Indels: 6
DB: 9 Gaps: 2

US-10-039-183A-4 (1-399) x US-09-815-242-3832 (1-1185)

QY      1 MetAlaLySGluValPheAsnArgThrLySP-ProHisValAsnileGlyThrileGlyHis 20
DB      1 ATGCGAAGAAAGAAATTTGACCGCTTCTTAATCCCATGTTTAACATTTGATTCGACAC 60
QY      21 Valaaphisgilylvalthrleuseralaalileseralavalleuserleuysgly 40
DB      61 GTTGACCATGTTAAACTATACATTAACAGCTGCAATTCGTATTAATCAAAACAGGT 120
QY      41 LeuAlaGluMeLysAspTyrAspAsnileAspAsnileAProGluGluLysGluArgGly 60
DB      121 GCGCGGGAAGCACAAAGCTACATTCATCGATTAACGCTCCAGAAAGAAAGAAAGCTGGA 180
QY      61 ILeThrileAlaThSerHileGlyLurGluThrGluAsnArgHileTyrAlaHileVal 80
DB      181 ATCACAACTACACTTCTCATATCGAATTAAGAACTGAACCTGCATATACACAGCTT 240
QY      81 AspCyProGlyHisAlaAspTyrValIlyAsnMetileThrGlyAlaAlaGluMetAsp 100
DB      241 GACTGCCACAGACACAGCGGACTAGCTTAATAAATGATGATGATGCTGCTCAATGAGC 300
QY      101 GlyAlaileuValvalserAlaAspGlyProMetProGlnThrArgGluHile 120
DB      301 GGAGCTATCTTAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY      121 LeuLeuserArgGluValGlyValProHisileValPheLeuAsnLysGluAspMet 140
DB      361 TTATTAATACCGTAACGTTGTGTATACATCATCGTTGATTTTAAACAAATGGAATAG 420
QY      141 ValaAspArgGluLyuLeuLyuLyuValGluMeGluValArgGluLeuLeuserAla 160
DB      421 GTTGATGACGAATATTCATTAATTAAGTAAGAAAGAAAGTTCGTAATTAATTAACAGA 480
QY      161 TyrGluPheProGlyAspAspThrProileValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB      481 TACGATTTCCCGAGCGATGATGTTCCAGTTATCGCAGGTTCTGCTTGAACCTTAAAGA 540
QY      181 GluAlaLybAlaGlyAsnValGlyLurProGlyGlyLyuValLeuLysLeuMetAlaGlu 200
DB      541 GCGCAGAGGTCT-----TATGAAGAAAGAAATTTGAATTAATGCTGCA 585
QY      201 ValaAspAlaTyrileProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
DB      586 GTTGACGAATATATCCCACTCCAGACGTGATACGAAACAACTTATGATGCGACG 645
QY      221 giuaapvalpaseserilealaaglyarglythrvalvalthrvalargylegulargly 240
DB      646 GAAGACGATTTCTCAATCACTGACGTGATGTTGCTACACAGACGCTTAAACGTGCT 705
QY      241 valvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalvalval 259
DB      706 GAAGTTGCGGTGTGACGAAGTTAAATCGTTGATTAATAAGCAAGAAACATCTTAACAA 765
QY      260 ThrValThrGlyValGluMeCpheArglysgluuvalvalvalvalvalvalvalvalval 279
DB      766 ACTGTTACAGGTGTTGAATGTTCCGTAATTAATTAAGTAAGTAAGTAAGTAAGTAAGTAAG 825
QY      280 ValGlyValleuLeuArgGlyThrLySlysgluGluValGluArgGlyMetValLeuCyS 299
DB      826 ATCGGTGCTTATTAATCGGTGTGATGACGATGAAGATATGAACTGAACTGAACTTAACT 885

```

Qy	300	YPSProGlySerIleThhProH1slyLysPheGlyIguIleItyrValIleuSerIys	319
Db	886	AAACCACTACATACATCTCCACACACAAATTCAAGCTGAAGTATACGTTATATCAAA	945
Qy	320	GIUGIUGIyGIyARH1sThhProHephThraenItyrArgProGInPheTyrValArg	339
Db	946	GAAGAGGCGGACGTCACATCTCATTTCTACTACCGCTCATTTCTACTTCGGT	1005
Qy	340	ThhIraPpValThhGlySerIleThhIreupProGlyIyValGInMetValMetProIly	359
Db	1006	ACACACAGCGTACTGCGHGTGTTGAATATGCCAGAGGACTGAATAGTATAGCTGGT	1065
Qy	360	AspAenValIyIleIthValGInIleuIleSerProValAlaIleGInIleuGlyThhIys	379
Db	1066	GATACGTTGCTATGACGCTTGAAATTAATTCACCCAAATCGCTATGAAGACGGAACCTGT	1125
Qy	380	PheAlaIleIeaArgIUGIyGIyARgThrValGlyAlaGlyValIseIraenIleIleuI	399
Db	1126	TTCTTATTCGTAAGGCGGACGTAAGTATAGGTCAGCGTGTATACGAATCGTTAA	1185

RESULT 8

```

? Sequence 6275. Application US/10282122A
? Publication No. US20040029123A1
? GENERAL INFORMATION:
? APPLICANT: Wang, Liangsu
? APPLICANT: Zamudio, Carlos
? APPLICANT: Malone, Cheryl
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlsen, Karl
? APPLICANT: Zyekind, Judith
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John
? APPLICANT: Carr, Grant
? APPLICANT: Yamamoto, Robert
? APPLICANT: Forsyth, R.
? APPLICANT: Xu, H.
? TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
? FILE REFERENCE: EPITRA.034A
? CURRENT APPLICATION NUMBER: US/10/282,122A
? PRIOR FILING DATE: 2003-02-20
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/230,335
? PRIOR FILING DATE: 2000-09-06
? PRIOR APPLICATION NUMBER: 60/230,347
? PRIOR FILING DATE: 2000-09-09
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/267,636
? PRIOR FILING DATE: 2001-02-09
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? Remaining Prior Application data removed - See File Wrapper or PALM.
? NUMBER OF SEQ ID NOS: 78614
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 6275
? LENGTH: 1185
? TYPE: DNA
? ORGANISM: Enterococcus faecalis
? US-10-282-122A-6275

```

[illegible]

```
Percent Similarity: 87.25% Conservative: 4
Best Local Similarity: 77.00% Mismatches: 6
Query Match: 78.71% Indels: 4
DB: 17 Gaps: 2
```

US-10-039-183A-4 (1-399) x US-10-282-122A-6275 (1-1185)

OY	MeAlLtyGsluLyAspSerhAqrThrLysPProHsValAaniLeGlyThrLeGlyHis	20
Db	1 ATGGCAAAAGAAAATTTTACCGGTTCTTAATCCCATGTGAACATTTGGTACATCGGACAC	60
OY	21 ValAspHisGclYLySerThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuYsgly	40
Db	61 GTTGACCATGTGTAAACTCATTTAACAGCTGCATTCCTACTGTATTATCAAAACACGGT	120
OY	41 LeuAlaGluMetLeuAspPyrTrpAspAsnIleAspAsnAlaProGluGluYsglyAsxGly	60
Db	121 GCGCGGAAAGACAAAGACTCAATTCATTCGATTAACGCTCCAGAAAGAAAAGAAACGTGGA	180
OY	61 IleThrIleAlaIleSerHisIleGluYrgIuThrGluAsnArgHisYrAlaHisVal	80
Db	181 ATCAACATCAACACCTTCTCATATCGAATATGTAAACTGAAACTCGTCACTATGACACGTT	240
OY	81 AspCysProGlyYHisAlaAspYrYVallyAsnMetIleThrGlyAlaAlaGluMetAsp	100
Db	241 GACTGGCCACGAGACATCCGACTCACTTAAAAACATGACATCGTGCTGCTCAAAATGGAC	300
OY	101 GlyAlaIleIleuValValSerAlaAlaAspGlyProMetProGluThrArgLysIle	120
Db	301 GGAGCTATCTTAGTAGTTTCTGCTCTGTATGCTTACTTACCTCAAAACAGTGAACATATC	360
OY	121 LeuLeuSerArgGluValGlyValProHisIleValValPheLeuAsnYsgluAspMet	140
Db	361 TTATTATTCACGTAAACGTTGGTGTACCAATACATCGTTGATTCTTAAAAAATGAGATATG	420
OY	141 ValAspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluValLeuLeuSerAla	160
Db	421 GTTGATGACGAAAGAAATTATTAGAAATTAGTAAAGAAATGGAAAGTTCGGTACTTATTCGAA	480
OY	161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu	180
Db	481 TTAGATTTCCACAGGCGATGATGTTCCAGTTATTCGACAGTTCTGCTTTGAAACGTTTGA	540
OY	181 GluAlaIlePheAlaGlyAsnValGlyGluTrpGlyGluYsglyValLeuYsgluMetAlaGlu	200
Db	541 GCGCAAGAGACT-----ATGAGAGAAAACCTTAGAATTAAATGGCTGCA	585
OY	201 ValAspAlaIleYrIleProThrProGluArgAspThrGluYsglyThrPheLeuMetProVal	220
Db	586 GTTGACGATATATATCCCACTCCAACTCAAGCTATCTGACAAACATTCATGATCCCAATC	645
OY	221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly	240
Db	646 GAAGAAGTATTCATCACTGAGAGGTGTACTGTGCTGACAGGCGCGTGTGAACGTGT	705
OY	241 ValValIleValValAspGluValGluIleValGlyIleArg---ProThrGluYsglyThr	259
Db	706 GAAGTTCGGGTGTGTCAGAACTTAATTCGTGTATTAAAGACGAAACATCTTAACACA	765
OY	260 ThrValThrGlyValAlaGluMetPheArgYsgluLeuGluYsglyGluAlaGlyAspAsn	279
Db	766 ACTGTTACAGGTGTGAATGTTCGTAAATTAATTAGACTACGCTGAAGCAGGCGAACAC	825
OY	280 ValGlyValLeuLeuArgGlyThrIleYsglyGluGluValAlaGluArgGlyMetValLeuCys	299
Db	826 ATCGGTGCTTATTACGTGTGTACAGTGAAGTAAATCGAATCGGACCAAGATTAATGCT	885
OY	300 LysProGlySerIleThrProHisYsglyPheGluGlyGluIleYrValLeuSerLys	319
Db	886 AAACCAAGCTACATCACTCCACACCAAAATTCAAAGCTGAAGTATACGTATTATCAAAA	945
OY	320 GluGluGlyValArgHisIleThrPhePheThrAsnYrArgProGluPheYrValArg	339

Db 946 GAAGAGCGGACGTACACATCTTCTACATACTACCTCTCAATTCTACTCTCGT 1005
Qy 340 ThrThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGly 359
Db 1006 ACAACAGACGTACGTGCTGTGTAGATTCAGAGAGGATGAGATGATGCTGCT 1065
Qy 360 AspaenValIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrIle 379
Db 1066 GATAACGTTGCTATGAGACGTTGAATTAATTCACCCAGTCTGATGAGAACGGAACCTGCT 1125
Qy 380 PheAlaIleArgGluGlyGlyValArgThrValGlyAlaGlyValIleSerAniIleIleGlu 399
Db 1126 TTCTCTATTCTGGAAGCGGACGTACTGAGTTCAGCGCTGTGTTACTGAATCGTTAA 1185

RESULT 9
US-09-815-242-6385
Sequence 6385, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyckind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6385
LENGTH: 1188
TYPE: DNA
ORGANISM: Enterococcus faecalis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1188)
US-09-815-242-6385

Alignment Scores:
Pred. No.: 4,92e-145 Length: 1188
Score: 1608.00 Matches: 308
Percent Similarity: 87.25% Conservative: 41
Best Local Similarity: 77.00% Mismatches: 45
Query Match: 78.71% Indels: 6
Gaps: 2

US-10-039-183A-4 (1-399) x US-09-815-242-6385 (1-1188)

Qy 1 MetAlaIleArgGluValPheAspaenThrIleProHISValAniIleGlyThrIleGlyHis 20
Db 1 ATGGCAAAAGAAATTTGACCGTTCTAATCCAGTTAACATTGATCGACAC 60
Qy 21 ValAspIleArgGluValThrLeuSerAlaAlaIleSerAlaValLeuSerLeuIleGlu 40
Db 1126 TTCTCTATTCTGGAAGCGGACGTACTGAGTTCAGCGCTGTGTTACTGAATCGTTAA 1185

Db 61 GTTGACATATGTAATACTAATTAACAGCTGCAATTCGTATTAATCAAAACAGCT 120
Qy 41 LeuAlaGluMetIleAspThrAspaenIleAspaenAlaProGluGluValArgGly 60
Db 121 GCGGGGAAAGCAAAAGCTACGATTCATCGTAACGCTCCAGAGAAAGAAAGAGCTGGA 180
Qy 61 IleThrIleAlaThrSerHisIleIleGluThrGluThrGluAspArgHisIleVal 80
Db 181 ATCAACATCAACACTTCATCATATGCAATATGAACTGAAACCTGCTACATGACACGCT 240
Qy 81 AspCysProGlyHisIleAlaAspThrValIleAspMetIleThrGlyAlaAlaGluMetAsp 100
Db 241 GACTGCCAGGACAGCGGACGTACCTGTAATAAATGATGATCACTGCTGCTCAAAATGAC 300
Qy 101 GlyAlaIleLeuValIleSerAlaAlaAspGlyProMetProGluThrArgIleHisIle 120
Db 301 GGAGCTATCTTAAGTATGTTCTGCTCTGATGCTGATGCTCAACAGCAACGTAACATATC 360
Qy 121 LeuLeuSerArgGluValIleValProHisIleValIlePheLeuAsnIleGlnAspMet 140
Db 361 TTATATATCAACGTAAGCTTGCTGTATACATACATCGTTGTTTAAACAAATGATATG 420
Qy 141 ValAspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
Db 421 GTTGATGACGAAGAATTTAGAAATTAGTAAGAAATGGAAGTTGAGACTTATATACAGA 480
Qy 161 TyrGluPheProGlyIleAspAspThrProIleValAlaGlySerAlaLeuArgAlaGlu 180
Db 481 TACGATTTCCAGGGAGATGATGTTCCAGTTACGAGGCTGCTTGAACCTTTAGAA 540
Qy 181 GluAlaIleValAlaGluValIleGluThrValIleValIleLeuIleLeuMetAlaGlu 200
Db 541 GCGGACGAGCTC-----TATGAGAAATAATTTGAATTAATGCTGCA 585
Qy 201 ValAspAlaIleProThrProGluArgAspThrGluValThrPheLeuMetProVal 220
Db 586 GTTGACGAATATATCCAACTCCAGACGTATGACCAACCAATTCATGATGACGATC 645
Qy 221 GluAspValPheSerIleAlaGlyArgGlyThrValIleThrGlyValArgIleGluArgGly 240
Db 646 GAAGACGTAATCTCAATCACTGAGAGCTGATCTGTTGCTACAGAGCGTGTAACTGAT 705
Qy 241 ValIleValIleValIleAspGluValGluIleValIleGlyIleArg---ProThrGluIleThr 259
Db 706 GAAGTTCGCGTTGAGACGAGACTTAATCGTTGATTAATAAGAGAAACATTAATAACA 765
Qy 260 ThrValIleThrGlyValGluMetPheArgIleGluLeuGluValGlyValIleAspAsn 279
Db 766 ACTGTTACAGGTTGGAATGTTCCGTAAATTAATTAAGACTACGTGAAGCGGACGACAC 825
Qy 280 ValGlyValIleLeuLeuArgGlyThrIleValGluGluValGluArgGlyIleValLeuCys 299
Db 826 ATCGGTGCTTAATACGTGCTGTAGACGTGAAGATGCAACGTGCAACATATTAAGCT 885
Qy 300 LysProGlySerIleThrProHisIleValIlePheGluGluIleThrValIleLeuSerIle 319
Db 886 AAACAGCTACATATCTCCACACAAATTCAAAGCTGAAGTATACCTATTATCAAAA 945
Qy 320 GluIleGluGlyIleArgHisIleThrProPhePheThrAsnIleThrArgProGluIlePheThrValArg 339
Db 946 GAAGAGCGGACGTACACACTTCATCTTCACTAATCCGCTCAATATTCATTCCTGCT 1005
Qy 340 ThrThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGly 359
Db 1006 ACAACAGACGTACGTGCTGTGTAGATTCAGAGAGGATGAGATGATGCTGCT 1065
Qy 360 AspaenValIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrIle 379
Db 1066 GATAACGTTGCTATGAGACGTTGAATTAATTCACCCAGTCTGATGAGAACGGAACCTGCT 1125
Qy 380 PheAlaIleArgGluGlyGlyValArgThrValGlyAlaGlyValIleSerAniIleIleGlu 399
Db 1126 TTCTCTATTCTGGAAGCGGACGTACTGAGTTCAGCGCTGTGTTACTGAATCGTTAA 1185

```
RESULT 10
US-10-282-122A-29444
; Sequence 29444, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29444
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-282-122A-29444

Alignment Scores:
Pred. No.: 6,83e-145      Length: 1185
Score: 1606.50           Matches: 302
Percent Similarity: 88.1% Conservative: 49
Best Local Similarity: 75.88% Mismatches: 42
Query Match: 78.63%      Indels: 5
DB: 17                  Gaps: 1
US-10-039-183a-4 (1-399) x US-10-282-122A-29444 (1-1185)

QY 1 MetAlaLySGluLySPheASnaRgThrLySProHISvalASnIleGIYThrIleGIYHis 20
DB 1 ATGGCTAAGGAAATTTGCAAGTGAACCAACCGCAAGTAAAGCTTGCGACCACTGCTAC 60
QY 21 ValAspHisGlyLyThrThreuseralaaIalieseralValleuserysgly 40
DB 61 GTTGACCATGTGTAACCAACCTGACTGCGCTTGACTACTATTGGCTAAAAAATTC 120
QY 41 LeuAlaGluMetLyAspLyThrAspAsnIleAspAsnAlaProGluGluLySGluArgly 60
DB 121 GCGGCTGCTGCAAAAGCTTACGACCAATCGACCAAGCCGCAAGAAAGAAACGACGCGGT 180
```

```
QY 61 IleThrIleAlaThrSerHisIleGIYThrGluThrGluAsnArgHisTyrAlaHisVal 80
DB 181 ATTACCATTAACACCTCCGACGTTGAATACGAAACCCGACCTACGACACGTA 240
QY 84 AspCysProGlyHisAlaAspTyrValLyAsnMetIleThrGlyValAlaGluMetAsp 100
DB 241 GACTGTCCGGGCGACCGCGACTGTAACCAATGATTAATCCGGCGCGACCAAAATGAC 300
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
DB 301 GGTGCAATCTCGTATGTTCCGACGCGACCGCCCTTAATCCGCAAAACCGCAACACATC 360
QY 121 LeuLeuseryArgGluValGlyValProHisIleValValPheLeuAsnLySGluAspMet 140
DB 361 CTGCTGCGCCGCTCAAGTAGGCGGTACCTACATCGTTTCATGAAACAAATGCGACATG 420
QY 141 ValAspAspGluGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
DB 421 GTCGACGATGCCGAGCTGTGGAACTGTGAATGAAATGAAATCCGCGACCTGCTCCAGC 480
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 481 TACGACTTCCCGCGGAGAGAGATGCCGATCGTACAGAGTTCCGACCTGAAGCCTTGAA 540
QY 181 GluAlaLyAlaGlyAsnValGlyGluThrGlyGlyValLeuLySGluMetAlaGlu 200
DB 541 GCGCATGCCGCT-----TACGAAGAAAAAATCTTCGAATTTGGCTGCTCA 585
QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLySerThrPheLeuMetProVal 220
DB 586 TTGACACGCTACATCCGACCTCCGAGCGCGCGTGGCAAACTTTCTTGTCCTATC 645
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
DB 646 GAAGACGATTTCTATTTCGGTGTGCTGACAGTACAGTACCGGTCGTAGAGCGCGT 705
QY 241 ValValLyValGlyAspGluValGluIleValGlyIleArgProThrGluLySerThr 260
DB 706 ATCATCCACGTCGGTGAAGAGATCGAATTCGTGCTGAAGAAACCTCAAAAACCACT 765
QY 261 ValThrGlyValGluMetPheArgGlyLeuGluGlyGlyGluAlaGlyAspAsnVal 280
DB 766 TGTACGGTGTGAATTTCCGAAACTGCTGGACGAAGGTCAAGCGACACACACGTA 825
QY 281 GlyValLeuLeuArgGlyThrLyLySGluGluValGluArgGlyMetValLeuCysHis 300
DB 826 GCGCATTTGCTGCGCGGTACCAACGTGAAGACGTAGAGCGTGTCAAGTATGCTAA 885
QY 301 ProGlySerIleThrProHisIleLySGluGlyGluIleTyrValLeuSerLySGlu 320
DB 886 CCGGCTACAAATCACTCTCACACCAAGTTCAAGACGAAGTATACGTGCTGACAAAGAA 945
QY 321 GluGlyLyArgHisIleThrProPheThrAsnTyrArgProGluPheTyrValArgThr 340
DB 946 GAGGCGGCGCGCAATCCCGTCTTCCCAACTACCGTCCCAATTTCACTTCCGTAC 1005
QY 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
DB 1006 ACCGACGTAACCGCGCGCGCTACTTTGGAAGAAGGTGGAATGTGTAATCGCGGCGAG 1065
QY 361 AsnValLySileThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLySph 380
DB 1066 AACGTAACCAATCAACGTAAGTGAATGTTCCGCTATCGCTATGGAAGAGTTTGCGCTT 1125
QY 381 AlaIleArgGluGlyLyArgThrValGlyValAlaGlyValLeuSerAsnIleIle 398
DB 1126 GCGATTTGCGAAGGCGCGGTACCGTGGTCCGCGGTGTTCTTCTGTATC 1179

RESULT 11
US-10-282-122A-29456
; Sequence 29456, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyn, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29456
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-282-122A-29456

Alignment Scores:
Pred. No.: 6,83e-145 Length: 1185
Score: 1606.50 Matches: 302
Percent Similarity: 88.19% Conservative: 49
Best Local Similarity: 75.88% Mismatches: 42
Query Match: 78.63% Indels: 5
DB: 17 Gaps: 1

US-10-039-183A-4 (1-399) x US-10-282-122A-29456 (1-1185)
QY 1 MetAlaValGluValPheAspArgThrIleValAsnIleGlyThrIleGlyHis 20
DB 1 ATGGCTAAGGAAAAATTGCAAGTAGCAAAACCGACGTAAAGTTGGACCACTCGCTAC 60
QY 21 ValAlaPheIleGlyValThrLeuSerAlaIleSerAlaValLeuSerLeuIleGly 40
DB 61 GTTGACATCGTAAACCAACCTGACTGCGCTTGACTGACTGACTGCTTAAATTC 120
QY 41 LeuAlaGluMetLeuAspTyrAspAsnIleAspAsnAlaProGluGluIleArgGly 60
DB 121 GCGCGTGTGCAAAAGCTTACGACCAATTCGACACGACCGAAGAAAGACACGCGT 180
QY 61 IleThrIleAlaThrSerHisIleGlyIleValThrGluAspArgHisIleTyrAlaHisVal 80
DB 181 ATTACCATTAACACCTCGACGTGAAATPACGAAACGAAACCGCACCTACGACACGTA 240
QY 81 AspCysProGlyHisAlaAspTyrValIleAspMetIleThrGlyAlaAlaGluMetAsp 100
```

```

DB 241 GACTGTCCGGGACACCGCACTACGTTAAACATGATTAACCGGCGACACAAATGAC 300
QY 101 GIYAlaIleLeuValIleSerAlaIleAspGlyProMetProGluThrArgGlyHisIle 120
DB 301 GGTGCAATCTGTGATGTTCGACACGACGCGCTTAATGCGCAAAACCGCAACACATC 360
QY 121 LeuLeuSerArgGluValAlaProHisIleValIlePheLeuAsnIleGlyIleAspMet 140
DB 361 CTGCTGCGCCGTCGAAGTAGGCGTTACTTACATCATCGTGTTCATGAACAAATGCCACATG 420
QY 141 ValAspAspGluIleLeuGluIleValIleGluMetGluValArgGluLeuLeuSerAla 160
DB 421 GTGCAAGATGCGGAGCTGTGGAAGCTGTAAGAAATCCGACCTGCTGTCCAC 480
QY 161 TyrGluPheProGlyIleAspAspThrProIleValAlaGlySerAlaLeuAspAlaGlu 180
DB 481 TACGACTTCCCGGCGGACGACTGCGGATCGTACAAAGTTCCGACATAAACCTTGGA 540
QY 181 GIYAlaValAlaGlyAsnValIleGlyIleProGlyGlyIleValIleLeuLeuMetAlaGlu 200
DB 541 GCGCATGCCGCT-----TACGAAGAAATCTTGAATTCGCTGCTGCA 585
QY 201 ValAspAlaTyrIleProThrProGluIleAspThrGluIleThrPheLeuMetProVal 220
DB 586 TTGCACTACATCATCCGACTCCGAGCGTCCGCGACCAAACTTCTGTTGCTATC 645
QY 221 GluAspValIlePheSerIleAlaGlyIleValIleThrValIleThrGlyIleGlyIleGly 240
DB 646 GAAGACGTAATCTCATTTCCGTCGTGTCAGTAGAACGCGTGTGATGAGCGCGT 705
QY 241 ValValIleValIleGlyAspGluValIleValIleGlyIleArgProThrGluIleThrThr 260
DB 706 ATCATCCACGTCGCGGACGATCAATTCGCGCTCGAAGAAATCTCAAAAAACCACT 765
QY 261 ValThrGlyValIleMetPheArgIleGluLeuGluIleGlyIleAlaGlyAspAsnVal 280
DB 766 TGTACCGGTGTGTAATGTTCCGCAAACTGCTGACGAAAGTCAGCGCGGCGACAAACGTA 825
QY 281 GlyValIleLeuAspArgGlyThrIleIleGlyIleGluValIleGluIleValIleLeuIle 300
DB 826 GCGGTATGCTGCGCGGTACCAACGTGAAGACGTGAACGCGGTGATTTGCGCAAA 885
QY 301 ProGlySerIleThrProHisIleIleAspIleGluIleIleThrValIleSerIleGly 320
DB 886 CCGGTACTATCATCTCTCACACCAATTCGAAGCAAGTACTGACTGACCAAGAA 945
QY 321 GluGlyIleArgHisIleThrPhePheThrAsnTyrArgProGluIleThrValIleArgThr 340
DB 946 GAGGCGCGCGGCATATCTCGTTCTTCCCAACTACCGTCCGCAATTCCTACTTCGCTAC 1005
QY 341 ThrAspValIleThrGlySerIleThrLeuProGluIleValIleGluMetValIleMetProGlyAsp 360
DB 1006 ACCGACGTACCGCGCGGTACTTGTGAAGAGGTGGAATGTGAATGCGGCGTGA 1065
QY 361 AsnValIleIleThrValIleGluLeuIleSerProValAlaLeuGluIleGlyThrIlePhe 380
DB 1066 AACGTAACTATTCGTTGAATGTAATTCGCGCTATACGCTATGGAAGAGCGTGGCTTT 1125
QY 381 AlaIleArgGluIleGlyIleArgThrValIleAlaGlyValIleSerAsnIleIle 398
DB 1126 GCGATTCCGAAAGGCGGTGTACCGTGGTGGCGGCGTGTCTTCTGTATTC 1179

RESULT 12
US-10-282-122A-29230
; Sequence 29230, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
```

```

/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See file wrapper or PAM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 29230
/ LENGTH: 1182
/ TYPE: DNA
/ ORGANISM: Neisseria gonorrhoeae
US-10-282-122A-29230

Alignment Scores:
Pred. No.: 4 01e-144 Length: 1182
Score: 1598.50 Matches: 300
Percent Similarity: 87.94% Conservative: 50
Best Local Similarity: 75.38% Mismatches: 43
Query Match: 78.24% Indels: 5
DB: 17 Gaps: 1

US-10-039-183A-4 (1-399) x US-10-282-122A-29230 (1-1182)
QY 1 MetAlaLySGluYbPheAsnArqThrLySProHisValAsnIleGlyThrIleGlyHis 20
DB 1 ATGGCTAAGGAAAAATTCACAGCTAGCAACCGACGTAAACGTTGGACCACTCGCTCAC 60
QY 21 ValAspHisGlyLySerThrLeuSerAlaAlaIleSerAlaValLeuSerLeuYsgly 40
DB 61 GTTGACCAATGTGTAAACCACTCGCTGCGCTTGAAGTACTATTATTGCTAAAAAATTC 120
QY 41 LeuAlaGluMetLyAspTyrAspAsnIleAspAsnAlaProGluGluYsgly 60
DB 121 GCGGCGCTGCAAAAGCTTACGACCAACCAACGACGACCCGAAAGAAAAACACGCGGT 180
QY 61 IleThrIleAlaThrSerHisIleGlyTyrGluThrGluAspArgHisTyrAlaHisVal 80
DB 181 ATTACCATTTAAACCTCGACCTAGCAATTCGAAACCGAAACCGCACCTACGACACGTA 240
QY 81 AspGlySPROGLYHISAlaSPTYrVallyAsnMetIleThrGlyAlaAlaGluMetAsp 100
DB 241 GACTGTCCGGGTACGCGCGACTACGTAAACATGATTAACGCGCGCCGCAAAATGAC 300
QY 101 GlyAlaIleLeuValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
DB 301 GGTGCAATTCGTGTATGTTCTGCTGCCGCGGCTTATGCGCAACCGCGCAACACATC 360

```

```

QY 121 LeuLeuSerArgGluValGlyValProHisIleValIlePheLeuAsnLySGluAspMet 140
DB 361 CTGCTGGCCCGGTCAAGTAGGCGGTACTTACATCATCGTGTTCATCAACAAATGACATG 420
QY 141 ValAspArgGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
DB 421 GTCCAGCATGCGCGACCTGTGGAACTGTGAAAGAAATCCGCGACCTGTGTCAGC 480
QY 161 TrrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 481 TACGACTTCCCGCGGACGACGACCTCCGATCGTACAGATTCCGACCTGAAGCTTGAA 540
QY 181 GluAlaLySAlaGlyAsnValGlyLutrpGlyGluYsValLeuLySLeuMetAlaGlu 200
DB 541 GCGCATGCGCT-----TACGAAGAAAAAATCTTGAACTGCTACCCCA 585
QY 201 ValAspAlaTrrIleProThrProGluThrGluAspThrGluYsThrPheLeuMetProVal 220
DB 586 TTGGACACTACATCCGACCTCCGAGCGTCCGCGGACMAACCATTCCTGCTGCTATC 645
QY 221 GluAspValIlePheSerIleAlaGlyYArgGlyThrValIleThrGlyYArgIleGluYArgGly 240
DB 646 GAAGACGTGTTCTCATTTCCGCGCGGTACCGTACCGGCGCTGTAGAGGAGGT 705
QY 241 ValValLySValGlyAspGluValGluIleValGlyIleArgProThrGlnLySThrThr 260
DB 706 ATCATCCAGTGTGGACGAGATGAAATCGTCCGTGAAAGAAACCAAAACCAACACC 765
QY 261 ValThrGlyValGluMetPheArgGlyLeuLeuGluYsglyGluAlaGlyAspAsnVal 280
DB 766 TGTACCGCGGTGAAATGTTCCGAAACCTGTGACGAGGACGAGCGCGGCGACAAACGTA 825
QY 281 GlyValLeuLeuArgGlyThrLySlySGluGluValGluYArgGlyMetValLeuCySly 300
DB 826 GCGATTTGCTCGCGGATACCAACGTGAAGACGTGAACCGGTCAAGTATGGCCAA 885
QY 301 ProGlySerIleThrProHisLySlyAspGluGlyGluIleTyrValLeuSerLySglu 320
DB 886 CCGGCTACTATCATCTCCACACCAAGTTCAAGCAAGATGTACGTATTGACAAAGAA 945
QY 321 GluGlyYValYArgHisThrProPhePheThrAsnTyrArgProGluInPheTyrValArgThr 340
DB 946 GAGGCGCGCGGCATACCCCGTTTTCGCCCAATACGTTCCCAATTCCTACTTCGTTACC 1005
QY 341 ThrAspValIleThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
DB 1006 ACTGACGTACCGCGCGGCTTACTTTGAAAAAGGTGTGAAATGTAAATGCGCGGTGAG 1065
QY 361 AsnValLySlyIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLySlyPhe 380
DB 1066 AACGTAAACATTACTGTGAACCTGATTCGCTTACGTATGAGAGAGTCTGCGCTTT 1125
QY 381 AlaIleArgGluGlyYValYArgThrValGlyValIleValSerAsnIleIle 398
DB 1126 GCGATTCCCGAAGGCGCGGATACCGTGGGTCCGCGGTTCCTGTGTATC 1179

RBSUT 13
US-10-282-122A-11657
/ Sequence 11657, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haeselbeck, Robert
/ APPLICANT: Olsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.

```

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11657
LENGTH: 1188
TYPE: DNA
ORGANISM: Burkholderia cepacia
US-10-282-122A-11657

Alignment Scores:
Pred. No.: 5,04e-144 Length: 1188
Score: 1597.50 Matches: 309
Percent Similarity: 86.50% Conservative: 37
Best Local Similarity: 77.25% Mismatches: 49
Query Match: 78.19% Indels: 5
Gaps: 2

US-10-039-183A-4 (1-399) x US-10-282-122A-11657 (1-1188)

Qy 1 MetAlaLySGuLyvPheAspThrProHISValAsnIleGlyThrIleGlyHis 20
Db 1 ATGGCCAAAGGAAATTTGAGCGGACCAAGCCGACGTGAAGTTGTCATGCTGAC 60
Qy 21 ValAspHisGlyLyvThrLeuSerAlaIleSerAlaValLeuSerLeuLygLy 40
Db 61 GTTACCAACGCGCAACACACCTGACGCGGACGATCACGACGTTCTACCAAGAAATTC 120
Qy 41 LeuAlaGluMetLyvAspTyTrAspAsnIleAspAsnAlaProGluGluLygLy 60
Db 121 GCGCGCAAGGACGCTACGACCAATGTCACGCGGACCGGAAAGGCGCGGAC 180
Qy 61 IleThrIleAlaThrSerHisIleGluTyTrGluThrGluAsnArgHisTyAlaHisVal 80
Db 181 ATCAAGATCAACACGCGACACGTCGATGCAAGAAAGGCTAACCGCCACTACGACACGTC 240
Qy 81 AspCyvProGlyHisAlaAspTyTrAlaLyvAsnMetIleThrGlyAlaAlaGluMetAsp 100
Db 241 GACTGCCCGGCGGACGCTGATATGTGAAGAAATGATCAAGCGCGCGGACGATGAC 300
Qy 101 GlyAlaIleLeuValAlaSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
Db 301 GCGCGATCTGCTGTTGCTCGGACACGACGCGCGGATGCGCAAAACGCTGACACATC 360
Qy 121 LeuLeuSerArgGluIleValGlyValProHisIleValIlePheLeuAsnLygGluAspMet 140
Db 361 CTGCTGGCGCGCTGATGCTGCTTCGATCAATCATCGTGTCTTAACAAGTCACATG 420
Qy 141 ValAspArgGluIleLeuGluIleuValAlaGluMetGluValArgGluLeuSerAla 160

Db 421 GTGACGACGCTGAACCTGCTGACGCTGTCGATGATGAAGTTCGGAACTGCTGCAAG 480
Qy 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 481 TACGATCTCCCGGGGACGACACGCGATCGTGAAGGTTGCGCAAGCTGCGTGGAA 540
Qy 181 GluAlaLySAlaGlyAsnValAlaGlyIleThrGlyGlyIleValLeuLeuMetAlaGlu 200
Db 541 -----GGGACACGCGGACGCTGCGGCGCAAGTGGCATATGACCTGCGCAC 588
Qy 201 ---ValAspAlaTyTrIleProThrProGluArgAspThrGluLyvThrPheLeuMetPro 219
Db 589 GCGCTGACACGATCATCCGACCGCGGACGTGACATGACGCGCGCTTCCTGATGCGG 648
Qy 220 ValGluAspValPheSerIleAlaGlyvArgGlyThrValIleThrGlyvArgIleGluArg 239
Db 649 GTGAGAGCGTGTCTCATCTCGGCGCGGTATGAGGTGAGCGGTCGTCTCAAGCGC 708
Qy 240 GlyValValLySValAlaAspGluValGluIleValGlyIleArgProThrGluLyvThr 259
Db 709 GGCATCGTGAAGCTGCGGACGAGAAATCGAATTCGTGATCAACCGACGCTGAAGACG 768
Qy 260 ThrValThrGlyValGluMetPheArgLySLeuGluIleValGlyAlaGlyAspAsn 279
Db 769 ACTGCAACGCGGCTGAATATGTTCCGCAAGCTGCTGACCAAGGTCAAGCGCGGACAAAC 828
Qy 280 ValGlyValIleLeuArgGlyIleThrLyvArgGluIleValGlyvArgIleValLeuCyS 299
Db 829 GTGGATCTCTGCTGCGCGGACCAAGGCTGAAGCGTGAAGCGTGGCAGGTTCTGGCG 888
Qy 300 LysProGlySerIleThrProHisIleLyvAspPheGluGluIleTyValLeuSerLyS 319
Db 889 AAGCGGTTGATCAACCCGACACGCACTTACCGCTGAAGTACGTGTCAGCAAG 948
Qy 320 GluGluGlyGlyvArgHisIleThrProPheThrAsnTyTrArgProGluPheTyValArg 339
Db 949 GACGAAGCGCGCGGTCACACCGCGTTCTTCAACAACCTACCGTCCGACATTTACTTCGT 1008
Qy 340 ThrThrAspValIleThrGlySerIleThrLeuProGluGlyValGluMetValMetProGly 359
Db 1009 ACGACGACGTCGACGCGCTCATCGACTGCGGAGCAAGAAATGTATGTCGCGAC 1068
Qy 360 AspAsnValLySileThrValGluLeuLeuSerProValAlaLeuGluLeuGlyThrLyS 379
Db 1069 GACAACTGTGATCACTGATGAACTGATCGCTCGATCGGATGAAGAGTCTGCGC 1128
Qy 380 PheAlaIleArgGluGlyGlyvArgThrValGlyAlaGlyValAlaSerAsnIleIleGlu 399
Db 1129 TTCGGATCCCGGAAGCGCGCTACGCTGCGCGCGCTGCTGCCAAGATCATCGAG 1188

RESULT 14
US-10-282-122A-14300
Sequence 14300, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Ilangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykend, Judith
APPLICANT: Wall, Daniel
APPLICANT: Twilick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21

```
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 14300
/ LENGTH: 1188
/ TYPE: DNA
/ ORGANISM: Burkholderia mallei
US-10-282-122A-14300

Alignment Scores:
Pred. No.: 1,536-143 Length: 1188
Score: 1592.50 Matches: 309
Percent Similarity: 86.00% Conservative: 35
Best Local Similarity: 77.25% Mismatches: 51
Query Match: 77.95% Indels: 5
DB: 17 Gaps: 2

US-10-039-183A-4 (1-399) x US-10-282-122A-14300 (1-1188)
QY 1 MetAlaLySGluYrPheAsnArgThrLySProHISValAsnIleGlyThrIleGlyHis 20
DB 1 ATGGCAAGAAAGAAATTTGAGCGGACCAAGCCGACGTGAAGCTTGATGATGCTGAC 60
QY 21 ValAspHisGlyYrThrIleuSerAlaIleSerAlaValIleuSerIleuGly 40
DB 61 GTTGACCAAGCGCAACGACGCTGACGCGACGATCGGACGCTGCTGCGGAAGTTC 120
QY 41 LeuAlaGluMetLyAspTyTrsAspAsnIleAspAsnAlaProGluGluYrGly 60
DB 121 GCGCGCAAGCGCAAGAGTACGACGAATCGACGCGCGCGCGGAAAGGCGCGGC 180
QY 61 IleThrIleAlaThrSerHisIleGlyIuTrGluAsnArgHisTyTrAlaHisVal 80
DB 181 ATCAAGATCAACACCGCCGACATCGATGCAAGAACGCGCAACCGCACTACCAACG 240
QY 81 AspCySProGlyHisAlaAspTyTrValIysAsnMetIleThrGlyAlaAlaGluMetAsp 100
DB 241 GACTCCCGCGGCGCAAGCGGACTGATGAAACATGATCAACGCGCGCGGACATGAGC 300
QY 101 GlyAlaIleLeuValValSerAlaIleAspGlyProMetProGluThrArgGluHisIle 120
DB 301 GCGCGATCTGTGTGCTCGCGCCGCTGACGCGCGGATGCGCAACCGGTGAGCAATC 360
QY 121 LeuIleuSerArgGluValAlGlyValProHisIleValAlaPheLeuAsnIleGluAspMet 140
DB 361 CTGCTGCGCGCGTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 141 ValAspAspGluGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
DB 421 GTGACGACGCGGAGACTGCTGAGCTGTGCGAAATGGAAGTGGGAACTGCTGCGAAG 480
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB 481 TACGACTTCGCGGCGGACGACGCGGATCATCAAGGCTTCGCGCAAGCTGCGCTGGA 540
```

```
QY 181 GluAlaValAlaGlyAsnValGlyGluTrpGlyGlu---LysValIleuLysMetAla 199
DB 541 -----GGGCAACAGGCGAGCTGGGCGCAAGTGGCGATCATCACTGCGCCGAC 588
QY 200 GluValAspAlaTyTrIleProThrProGluLysAspThrGluLysThrPheLeuMetPro 219
DB 589 GCGCTGACACGTCATCTCCACGCGGAGGCTGGCTGACGCGGCGCTCTGATGTCG 648
QY 220 ValGluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArg 239
DB 649 GTGAAAGCTGTCTTCATCTCGGCGCTGATGATGATGATGATGATGATGATGATG 708
QY 240 GlyValValLysValGlyAspGluValGluIleValGlyIleArgProThrGluLysThr 259
DB 709 GCGGTGATCAAGGTTGGCGGAGAAATCGAAATCGGTGATCAAGCGGACGCGAAGCAG 768
QY 260 ThrValThrGlyValGluMetPheArgGlySerGluLeuGluYrGlyGluAlaGlyAspAsn 279
DB 769 ACTGCAAGGCGCGGAAATGTTCCGCAAGCTGTGACCAAGGTCAGGCTCAGGCGGAC 828
QY 280 ValGlyValIleuLeuArgGlyThrLySbysGluGluValGluArgGlyMetValLeuCyS 299
DB 829 GTGCTATCTGCTGCTGCGCGGACCAAGCGTGAAGCGTGAAGCGCGCGGACGTTTGGCG 888
QY 300 LysProGlySerIleThrProHisLysLysPheGluGluGlyIleTyTrValIleuSerLys 319
DB 889 AAGCGGCTGATCGATCAGCGCCGACGCACTTCAAGGCAAGTGTACGTGCTGAGCAAG 948
QY 320 GluGluGlyGlyAlaGlyHisThrProPhePheThrAsnTyTrAlaGlyProGluPheTyTrAlaArg 339
DB 949 GACGAAAGCGCGCGGACACCGCGTCTTCAACAACTACCGTCCGACGTTCTACTTCGCT 1008
QY 340 ThrThrAspValIleThrGlySerIleThrIleuProGluGlyValGluMetValMetProGly 359
DB 1009 ACGACGCGTGAAGCGGCTTCATGCACTGCGCGGAAAGCAAGAAATGTATGCGCGGC 1068
QY 360 AspAsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLys 379
DB 1069 GACAAAGTGTGATCAACGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1128
QY 380 PheAlaIleArgGluGlyValArgThrValAlaGlyValValSerAsnIleIleGlu 399
DB 1129 TTCGATCTCCGCAAGGCGGTGCGACCGTCCGCGCGGCTGCTGCGCAAGTCAATCGAG 1188

RESULT 15
US-10-398-221-9
/ Sequence 9, Application US/10398221
/ Publication No. US2004001851A1
/ GENERAL INFORMATION:
/ APPLICANT: KUNST, Frederik
/ TITLE OF INVENTION: Listeria innocua, genome and applications
/ FILE REFERENCE: 344 702 - US
/ CURRENT APPLICATION NUMBER: US/10/398, 221
/ PRIOR FILING DATE: 2003-03-27
/ PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
/ PRIOR FILING DATE: 2001-10-04
/ PRIOR APPLICATION NUMBER: FR 00/12 697
/ NUMBER OF SEQ ID NOS: 4025
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 9
/ LENGTH: 684707
/ TYPE: DNA
/ ORGANISM: Listeria innocua
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1)..(end)
/ OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

Alignment Scores:
Pred. No.: 4,866-140 Length: 684707
```


Score: 1592.00 Matches: 297
 Percent Similarity: 87.19% Conservative: 50
 Best Local Similarity: 74.62% Mismatches: 45
 Query Match: 77.92% Indels: 6
 DB: 17 Gaps: 2

US-10-039-183a-4 (1-399) x US-10-398-221-9 (1-684707)

QY 1 MetAlaValGluValPheAsnArgThrIleProHisValAsnIleGlyThrIleGlyHis 20
 Db 462669 ATGGCAAGAAAGAAATTTGACCGCTCTAAACCCCATGTTAACATTGGTACTATTGGACAC 462728
 QY 21 ValAspHisGlyLeuThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuGly 40
 Db 462729 GTTGACCATGTGAAACCTACTTAACTGACGTGAATTACACAGTACTTCTGCTAAAGAGC 462788
 QY 41 LeuAlaGluMetLeuAspTyrAspAsnIleAspAsnAlaProGluGluValArgGly 60
 Db 462789 TTTGCTGATGCAACAGCTTATGATCAAAATTGATGCTGCCAGAAAGAAAGAACGTGCT 462848
 QY 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAspArgHisTyrAlaHisVal 80
 Db 462849 ATCAAACTCTACTGCTCAGCTGATGATGACCAACTGACCAACCGTCACTATGACACAGTT 462908
 QY 81 AspCysProGlyHisAlaAspTyrValIleAspMetIleThrGlyAlaAlaGluMetAsp 100
 Db 462909 GACTGCCAGAGACATGCCGATTAACCTTAAACATGATCATCTGCTGACCAAAATGAC 462968
 QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgGluHisIle 120
 Db 462969 GGAGCTATCTTAGTAGTAGTATCTGCTGCTGATGCCCAATGCCCAAACTCGTGAACATATC 463028
 QY 121 LeuLeuSerAlaArgGlnValGlyValProHisIleValPheLeuAsnLysGluAspMet 140
 Db 463029 TTACTTTCACGTCAGAGTGGTGGTTCATACATCGTTGATTCATGACAAATGTGCATG 463088
 QY 141 ValAspAspGlnGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
 Db 463089 GTTGACGATGAAAGATTACTAGCAATTAGTGAATGCAATTCGTGATCTATTAACTGAA 463148
 QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
 Db 463149 TATGAATTCCTCGCGATGATCATCTGTAATCAAAAGTTGAGTCTTAAAGCACTTCA 463208
 QY 181 GluAlaLysAlaGlyAsnValGlyLysIleTyrGlyLysValLeuLysLeuMetAlaGlu 200
 Db 463209 -----GGTGAAGCTGATCGGAAAGCTAAATTAATGAGAGCT 463253
 QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
 Db 463254 GTAGATTCTTACATTCCAACTCCAGAACGTGATCTGACAAACCAATTCATGATGCCAGTT 463313
 QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrValArgIleGluArgGly 240
 Db 463314 GAGGATGATTCCTCAATCACTGCTGCGAACAAGTTCGCACTGCAAGCTGGAACGTGGA 463373
 QY 241 ValValLysValGlyAspGluValGluIleValGlyIle---ArgProThrGlnLysThr 259
 Db 463374 CAAGTTAAAGTTGGTGAAGAGTAGAAGTATCGGATTTGAAAGAAAGCAAAAGTA 463433
 QY 260 ThrValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsn 279
 Db 463434 GTAGTAAGTGAAGTAAGTAAGTTCGTAATTAATTAAGTACAGTGAAGCTGGCAAC 463493
 QY 280 ValGlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCys 299
 Db 463494 ATGGCGCACTTCTACGTGGTCTGCTGTAAGATATCCAAAGTGTCAAGTATTAAGCT 463553
 QY 300 LysProGlySerIleThrProHisLysLysPheGluGlyLysIleTyrValLeuSerLys 319
 Db 463554 AAACCAAGGTTCATTAATCAACACTTAAGCTGAAGCTTAATGTTTAACTAA 463613
 QY 320 GluGluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArg 339

Db 463614 GAAGAAAGTGAACGTCACATCTCATTTCAACAACTACCGCCCAAAATTTATTTCCGT 463673
 QY 340 ThrThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGly 359
 Db 463674 ACTACTGACGTAACTGATTTGTATACCTTCAGAAAGGTACTGAATGTAATGCTTGT 463733
 QY 360 AspAsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLys 379
 Db 463734 GATTAATTAAGCTTGCAGTTGAATTAATTCACCAATCCCTATGCAAGACGTAATAA 463793
 QY 380 PheAlaIleArgGluGlyGlyArgThrValGlyAlaGlyValSerAsnIle 397
 Db 463794 TTCTCTATCCGTGAAGGCGGAGGTACAGTAAAGCGCTGCGTGTGTTTAAACATC 463847

Search completed: April 24, 2005, 12:12:39
 Job time : 1087.99 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 24, 2005, 05:28:36 ; Search time 4475.89 Seconds
(without alignments)
3393.213 Million cell updates/sec

Title: US-10-039-183a-4
Perfect score: 2043
Sequence: 1 MAKEKFNTRKPHVNIPTIGH.....FAIREGRTVGAGVSNITIE 399

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US10039183/runat_22042005_122425_20124/app_query.fasta_1.1038
-DB=EST-QFMT=fastap -SUPFIX=est -MINMATCH=0.1 -LOOPEXT=50
-UNITS=bites -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10039183_@CGN_1_1_5533@runat_22042005_122425_20124 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLDOP=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: gb_esc1.*
2: gb_esc2.*
3: gb_esc3.*
4: gb_esc4.*
5: gb_esc5.*
6: gb_esc6.*
7: gb_esc7.*
8: gb_esc8.*
9: gb_esc9.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1436	70.3	1617	3	CNS0A2SS
2	1432	70.1	1770	3	CL966994 Arabidops
3	1431	70.0	1594	3	CNS0A3M3
4	1416.5	69.3	1815	3	AY104241 Zea maye
5	1365	66.8	1575	3	CNS0A3N6
6	1227.5	60.1	1611	3	AK075857 Arabidops
7	1227.5	60.1	1629	3	AK084724 Mus muscu
8	1226.5	60.0	1494	3	CR593603 full-leng
9	1226.5	60.0	1505	3	CR621980 full-leng

10	1226.5	60.0	1534	3	CR608035 full-leng
11	1095	53.6	1119	3	AY105509 Zea maye
12	1094	53.5	1193	3	CK215616 FCAS02758
13	1064	52.1	960	9	CNS06YH3
14	1063.5	52.1	899	7	CO083417 GR_Ea48H
15	1057.5	51.8	879	7	CV291218 Arabidops
16	1043	51.1	922	9	CG155224 OXKAU17TV
17	1030.5	50.4	868	7	CG130623 GR_EB33H
18	996.5	48.8	861	6	CB981675 Arabidops
19	989.5	48.4	942	7	CK178095 Arabidops
20	987.5	48.3	802	7	CF444610 Arabidops
21	985	48.2	893	3	CO009744 Arabidops
22	982.5	48.1	805	6	CD818818 Arabidops
23	972.5	47.6	803	7	CO127418 Arabidops
24	963	47.1	993	7	CO011949 Arabidops
25	959.5	47.0	909	7	CF355578 Arabidops
26	957.5	46.9	822	7	CF393970 Arabidops
27	957	46.8	894	7	CK188451 Arabidops
28	955.5	46.8	823	6	CB652301 Arabidops
29	955	46.7	954	7	CF700090 Arabidops
30	948	46.4	882	7	CK178021 Arabidops
31	944	46.2	868	7	CO083742 Arabidops
32	939.5	46.0	774	6	CA916852 Arabidops
33	933.5	45.7	1018	1	AL556182 Arabidops
34	930	45.5	929	7	CF686795 Arabidops
35	924.5	45.3	969	7	CK286254 Arabidops
36	919.5	45.0	783	9	CG358260 Arabidops
37	918.5	45.0	758	4	BG887215 Arabidops
38	913.5	44.7	1138	8	BZ549679 Arabidops
39	908	44.4	825	6	CB644718 Arabidops
40	907	44.4	816	7	CO110658 Arabidops
41	906.5	44.4	706	7	CF855315 Arabidops
42	905.5	44.3	947	5	BX839381 Arabidops
43	901.5	44.1	896	7	CF584519 Arabidops
44	900.5	44.1	765	6	CB654737 Arabidops
45	898	44.0	1106	7	CK211377 Arabidops

ALIGNMENTS

RESULT 1
CNS0A2SS
LOCUS
DEFINITION
CNS0A2SS
1617 bp mRNA linear HTC 06-FEB-2004
Arabidopsis thaliana Full-length cDNA Complete sequence from clone GS11562202 of Adult vegetative tissue of strain col-0 of Arabidopsis thaliana (thale cress).

ACCESSION
BX827505
VERSION
BX827505.1 GI:42460514
KEYWORDS
HTC; GBLT cDNA.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE
1 (bases 1 to 1617)
Castelli V., Aury J.M., Jallion O., Winkler P., Clepet C.,

AUTHORS
Menard M., Crnaud C., Quetier F., Scarpetti C., Schachter V., Temple G., Caboche M., Weissenbach J. and Sallanoubat M.
TITLE
Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation

JOURNAL
REFERENCE
Unpublished
2 (bases 1 to 1617)
Genoscope.

COMMENT
Direct Submission
Submitted (18-NOV-2003) Genoscope - Centre National de Séquencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
The sequences are based on single pass reads.
Life technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jallion O., Winkler P., Menard M., Crnaud C.,

Schachter V., Weissenbach J., Salanoubat M.
 URV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis
 genome released by MIPS (Munich Information center for Protein
 Sequences). 5 prime and 3 prime are assembled with Phrap.
 http://www.genoscope.cns.fr/externe/sequences/banque_Projet_EF/full
 length
 http://www.genoscope.cns.fr/cgi-bin/gsb/gsb?source=Arabidopsis.
 Location/Qualifiers

FEATURES

source

1. 1617
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Col-0"
 /db_xref="taxon:3702"
 /clone="GSLT8662D02"
 /tissue_type="adult vegetative tissue"
 /plasmid="pCMVSPORT_6"
 1. 1617
 /gene="At4g20360"

ORIGIN

Alignment Scores:

Pred. No.: 1,44e-149 Length: 1617
 Score: 1436.00 Matches: 280
 Percent Similarity: 80.44% Conservative: 49
 Best Local Similarity: 68.46% Mismatches: 68
 Query Match: 70.29% Indels: 12
 Gaps: 3

US-10-039-183a-4 (1-399) x CNS0A2S5 (1-1617)

QY 2 AAlaYsGluYsPheAsnArGThrLysPProHisValAsnIleGlyThrIleGlyHisVal 21
 Db 250 GCTCGTGAAGAAATTCGAGAGAGAAAGCCCTCATGTCACATCGAACCCTCGCTCATGT 309
 QY 22 AspHisGlyLysThrThreUsenSerAlaIleSerAlaValLeuSerLeuLysGlyLeu 41
 Db 310 GACCATGGGAAATCTTAACCGCAGCTTAACCATGCTCGCTCCATGCTGCTTCC 369
 QY 42 AlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGlyIle 61
 Db 370 AGCGTCGCTAAAGACACACAGATGACGCTGCGCCGAGAGAGAGACCTGCTGATAC 429
 QY 62 ThrIleAlaThrSerHisIleGlyLysThrGluThrGluAsnArgHisTyrAlaHisValAsp 81
 Db 430 ACAATCAACACCTGCTACTGTTAGTACGAGACTGAGAACTGTCACCTACCTCATGTAT 489
 QY 82 CysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGlnMetAspGly 101
 Db 490 TGTCTCGTCACGCGATTAACGTTAAGATATGATTACCGAGCTGACAGATGACGGA 549
 QY 102 AlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGlnHisIleLeu 121
 Db 550 GGTATCTCTCGTGTTCGCGCCGACGATGCTTATGCTTCACTCACTAAAGACATATCTT 609
 QY 122 LeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMetVal 141
 Db 610 TTGGCTAAGCAGCTGGTGTCTCTGATATGTTGTTGTTCTTAAACAAGAGATCAAGTA 669
 QY 142 AspAspGlnGluLeuGluGluLeuValGluMetGluValArgGluLeuLeuSerAlaTyr 161
 Db 670 GATGATGACGAGATTCTAGAGCTCGTTGAGCTTGAGGTTCCGAGCTTCTCGCTTAT 729
 QY 162 GluPheProGlyLysAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
 Db 730 GAATTTGACCGGATGATATTCGATTAATCTCTGCTTGTCTTTAGCCCTTGAAGCT 789
 QY 181 -----GluAlaLysAlaGlyAsnValGlyGluThrGlyGluLysValLeu 195
 Db 790 CTTACTGAGAAATCCTAAGGTTAAGAGAGTAT---AACAAATGGGTAGATTAAGATTAT 846
 QY 196 LysLeuMetAlaGluValAspAlaTyrIleProThrProGluArgAspThrGluLysThr 215

Db 847 GAACCTTAGATGCTGTGATGATTAATCCCTATCCCTCAGAGACAAATGATTCGA 906
 QY 216 PheLeuMetProValGluAspValPheSerIleAlaGlyArgGlyThrValThrGly 235
 Db 907 TTTCTGTAGCTGTTGAGAGATGTCTTCTATCACTGACGATGTAAGCTGCTACAGG 966
 QY 236 ArgIleGluArgGlyValValLysValGlyAspGluValGluIleValGlyIleArgPro 255
 Db 967 CGTGCAGAGAGGATGACGTTAAGGTAAGAGACGTAAGATTATGAGGTTTGAAGGAG 1026
 QY 256 ThrGlnLysThrThrValThrGlyValGluMetPheArgGlyGluLeuGluLysGly 275
 Db 1027 ACTAGAGCTTACACTGCTACTCGGGGTAAATGTTTCAGAAATTTGATGATGAGCTTTA 1086
 QY 276 AlaGlyAspAsnValGlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGly 295
 Db 1087 GCTGATGACAAATGTAAGGTTGTTGCTTGGGGTATTCAAAAGCGTGAATTTACAGAGCT 1146
 QY 296 MetValLeuCysLysProGlySerIleThrProHisLysLysPheGluGlyIleTyr 315
 Db 1147 ATGTTTAGCTAAGCCGGATCGATTAATCTCAATCAACAGTTTGAAGCAATTATCTAT 1206
 QY 316 ValLeuSerLysGluGluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGln 335
 Db 1207 GTGTTGAAGAAAGAGAGAGGTTGAGGCAATTCCTCATCTTGTGACGGGTACAGGCTCAG 1266
 QY 336 PheTyrValArgThrThrAspValThrGlySerIleThr-----LeuPro 350
 Db 1267 TTTCAATGAGAGCACTGATGTTTACGGGTAAAGTACAGAAAGATATGACAGCAAAAGT 1326
 QY 351 GluGlyValGluMetValMetProGlyAspAsnValLysIleThrValGluLeuIleSer 370
 Db 1327 GAAGAGTCGAAGATGTTATGATCCCGGTATGAGTGAAGTGAATGTTGTTGAGCTTATG 1386
 QY 371 ProValAlaLeuGluLeuGlyThrLysPheAlaIleArgGluGlyArgThrValGly 390
 Db 1387 CCGGTCCTTGTGAACAAAGGATGAGTTTCTTCTATCAGAGAGAGAAAGCTGTGTGT 1446
 QY 391 AlaGlyValValSerAsnIleGlyLeu 399
 Db 1447 GCTGAGTATATGGACATCTCTGAA 1473

RESULT 2
 CL966994 1770 bp DNA linear GSS 21-SEP-2004
 LOCUS OsIFCC014542 Oryza sativa sativa Expressed library Oryza sativa (indica
 DEFINITION OsIFCC014542 Oryza sativa sativa Expressed library Oryza sativa (indica
 ACCESSION CL966994
 VERSION CL966994.1 GI:52388637
 SOURCE GSS.
 ORGANISM Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eubharoidae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 1770)
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Zhao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G.K.S., Deng, X.W. and Wang, U.
 An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 Unpublished (2004)
 CONTACT Chen Chen
 JOURNAL Department of Bioinformatic
 COMMENT Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 FEATURES
 source 1. 1770
 Location/Qualifiers


```

/strain="COL-0"
/db_xref="Eaxon:3702"
/clone="GSLT8252B09"
/tissue_type="Adult vegetative tissue"
/plasmid="pCMSPORT 6"
/complement(1..1594)
/gene="At4g02930"

```

ORIGIN

Alignment Scores:

```

Pred. No.:      5,12e-149      Length:      1594
Score:          1431.00        Matches:      276
Percent Similarity: 82.12%      Conservative: 50
Best Local Similarity: 69.52%      Mismatches: 63
Query Match:    70.04%         Indels:      8
DB:             3              Gaps:         3

```

US-10-039-183a-4 (1-399) x CMS0A3M3 (1-1594)

```

QY 6 PheAenATgThrLysProHleValAsnileGlyThrileGlyHleValAspHleGlyLys 25
DB TTTACTCGAAATAAACCTCATGTAATGTGGAACATTTGGSCATGTCATATGCGCAAG 261
QY 26 ThrThrLeuSerAlaAlaileSerAlaValleuSerLeuLysGlyLeuAlaGluMetLys 45
DB ACCACTTTAACTGCTGCATACAAAGGCTTCTGCTGAGAGGCGCAAGACTTAAT 321
QY 46 AspTyArpAsnileAspAsnAlaProGluGluLysGluArgGlyLleThrileAlaThr 65
DB GCCTTGATGAATGATGAAGCTCTCGTAAGAAAGAGAGGAATTAATTAATTCACAG 381
QY 66 SerHileGlyLysGlyThrGluAsnArgHleGlyAlaHleValAspCysProGlyHle 85
DB GCTCATGAGGATAGAGCTCAAAAGGCTCACTATGCTCATGTCATGTCATGTCGACAC 441
QY 86 AlaAspTyValLysAsnMetileThrGlyAlaAlaGluMetAspGlyAlaileLeuVal 105
DB GCAGATTAATGTTAAGATATGATTAATCTGAGCTGCGCAATGATGCGCAATTCCTGG 501
QY 106 ValSerAlaAlaAspGlyProMetProGlnThrArgGluHleileLeuSerArgGln 125
DB GTTTGAGACCAAGATGAGCCCATGCGGACACAAAGAAACATATTTCTTCTGACGCCAG 561
QY 126 ValGlyValProHileValValPheleuAsnLysGlnAspMetValAspAspGlnGlu 145
DB GTTGGTGTTCCTCCTCATGTCGCTTCTTGAACAAAGTTGATGATGATACCTGAG 621
QY 146 LeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAlaTyGlnPheProGly 165
DB CTCTTGAGAGCTTGTGAGATGAGACTAGCTGAGCTCTCAGCTTCTACAAAGTTTCTGAG 681
QY 166 AspAspThrProileValAlaGlySerAlaLeuArgAlaLeuGluGluAlaAlaGly 185
DB GATGATTAATTCATATCCGAGAACTGCTGCTGCGCAATTAAG-----GAG 729
QY 186 AsnValGlyGluTrpGly---GluLysValleuLysLeuMetAlaGluValAspAlaTy 204
DB ACCAATGATGAATTTGGAAGCAAGCTATATTAAGCTGATGATGATGATGATGAT 789
QY 205 IleProThrProGluLysArgAspThrGluLysThrPheLeuMetProValGluAspValPhe 224
DB ATACCTGACCTGTCGCGCTCTGACAAAGCTTCTTGAATCCCAATGAAGATGTTTC 849
QY 225 SerileAlaGlyArgGlyThrValValThrGlyArgGlyleGluArgGlyValValysVal 244
DB TCAATTTCAAGAGCGTGAACCTGTTCAACCGGCTGATCGAAGCAAGGATCTTAAAGG 909
QY 245 GlyAspGluValGluileValGlyLleArgPro-----ThrgLileThrThrVal 261
DB GGTGAAGAGTTGATATGAGTTTACGTGAGGGGGGATTTCCACATGAATGCACTGTA 969
QY 262 ThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnValGly 281

```

```

DB 970 ACTGGGTTTGAGAGTTCCAGAAAGATTTTGATTAATGACAGCGCTGATTAATGTAGCA 1029
QY 282 ValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLysPro 301
DB CTCTTCTGCGTGGCTTAAAGAGAGACATTCAGCTGATGATGATGATGATGATGATGAT 1089
QY 302 GlySerileThrProHileLysLysPheGluGluGlyLleTyValleuSerLysGluGlu 321
DB GATTATCATCAAGACATCAAGAAAGTTTGAACAGAGATTTTACGTCTCACAAGAGATGAA 1149
QY 322 GlyGlyArgHleThrProPhePheThrAsnTyTrpProGlnPheTyValArgThrThr 341
DB 1150 GGTGACGCTGCACATGATTTTCTTACCTACAGGCTTCAGTTTACTTGAGAGCTGCA 1209
QY 342 AspValThrGlySerileThrLeuProGluGluValGluMetValMetProGlyAspAsn 361
DB CATATCACTGGCAAGTGAATTAATCCGAAACGTAAGATGTTATTCCTGTCATCAAT 1269
QY 362 ValLysileThrValGluLleuLleSerProValAlaLeuGluLeuGlyThrLysPheAla 381
DB GTCAACAGCTGTTTTCGAGTTATCATGCTGCTGCCACTCGAAACAGTCMAAGATTGCC 1329
QY 382 IleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAlailelle 398
DB TTAAGGGAAGAGAGTGAACAGTTGAGCTGCTGTTGATCAAAAGTGAATG 1380
QY 1330 TTAAGGGAAGAGAGTGAACAGTTGAGCTGCTGTTGATCAAAAGTGAATG 1380

RESULT 4
AY104241 1815 bp mRNA linear HTC 16-OCT-2002
LOCUS Zea mays PC0131873 mRNA sequence.
DEFINITION Zea mays AY104241
ACCESSION AY104241
VERSION AY104241.1 GI:21207319
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1815)
REFERENCE
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 1815)
Coe,E.H.
JOURNAL
Direct Submission
TITLE
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
COMMENT
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSU, maizeMap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
FEATURES
source
1..1815
location/Qualifiers
1..1815
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:637696"
/db_xref="Eaxon:4577"
/clone_lib="Maize Mapping Project/Dupont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
configs to seed Dupont configs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
ORIGIN
Alignment Scores:

```


Percent Similarity:	80.35%	Conservative:	45%
Best Local Similarity:	68.01%	Mismatches:	70%
Query Match:	66.81%	Indels:	9
DB:	3	Gaps:	3

US-10-039-183A-4 (1-399) X CNS0A3N6 (1-1575)

[illegible]

Db	1134	GGTGGAGGTCAACATGCACCTTCTCTCAACTACAGAGGCTCAGTTTACTTGAGGACTGCA	1193
Qy	342	AspValIthrGlySerIleIerhLeuProGluGlyValGluMetValMetProGlyAspAsn	361
Db	1194	GATTCACCTCGCAAGAGGGAATTATCCCGAAACGTAAGATGTTAAGCCTGGTGAACAT	1253
Qy	362	ValIysIlethrValGluIleuIleSerProValAlaLeuGluIleuGlyIThrIysPheAla	381
Db	1254	GTCACAGCTGTTTTCGGATTATATCATCTCTGTCCACTTCGAAACAGGTCAAAAGATTGCC	1313
Qy	382	IleArgIleuGlyGlyArgThrValGlyAlaGlyValIserAsnIleIle	398
Db	1314	TTAAGGAGAGAGGTAGAACAGTTGGAGCTGGTGTGTATCAAAAGTAGAG	1364
RESULT 6			
AK075857		1611 bp	mRNA
LOCUS			
DEFINITION			
AK075857			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			

Db 1284 ACCTGAGGATGCAACACCATGTGGCTGCTTGTATGATGTA 1334

RESULT 7
AK084724
LOCUS
DEFINITION
AK084724 1629 bp mRNA linear HNC 03-APR-2004
Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched library, clone:D3J0035P20 product:ELONGATION FACTOR TU,
MITOCHONDRIAL PRECURSOR (P43) homolog [Homo sapiens], full insert sequence.

ACCESSION
AK084724
VERSION
AK084724.1 GI:26351252
KEYWORDS
HNC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636

REFERENCE
AUTHORS
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159

REFERENCE
AUTHORS
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yonikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20350913
PUBMED
11076861

REFERENCE
AUTHORS
4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409, 685-690 (2001)
REFERENCE
AUTHORS
5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
REFERENCE
AUTHORS
6 (bases 1 to 1629)
Adechi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kaubawa, T., Kato, H., Kawai, O., Kojima, Y., Kondo, S., Kono, H., Konda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, P., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE
Direct Submission
JOURNAL
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Sueni-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,

COMMENT
URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>.
Location/Qualifiers
1. 1629
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM:DB:D3J0035P20"
/db_xref="taxon:10090"
/clone="D3J0035P20"
/tissue_type="heart"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
75. 1433
/note="unnamed protein product; ELONGATION FACTOR TU, MITOCHONDRIAL PRECURSOR (P43) homolog [Homo sapiens] (SWISSPROT P49411, evidence: FASTA, 93.8%ID, 99.7%length, match=1353)
putative"
/codon_start=1
/protein_id="BAC39263.1"
/db_xref="GI:26351253"
/translation="MAAATLTATPRTSGLCASPTPLQGLRPIKAPASFLRCGLA
VEAKTYVRDPRHVNVTGTLTAVDGLTTLTAIRKILAEAGGARFKYKEEDNPEER
ARGITLNAHAEYSATLAHVAHTPCGADYVAKMISTGATLDCILVAANDGPMQ
TREHLIAKOIGVHVVVYVYKAAVQSEVVELREILRELFEGYKGETPIYVG
SALCALRDRDEPLKSVQKLDVAVPTIIPYPTDLDKPELIPVESYVIGRGVVT
GTLERGLKKDBEELGHNKNITVTVTGIMFPHKSLERAAAGNLALVYKQRED
RRGVVWPVPSIOPQKVEAQVYLLSKESGRHPPVSHFMPVMSLTWMACTILP
PKKELAMGEBDLKSLIRQPMILEKGRFTLRDGNKTIIGTLVTDVPMTEBDKNIK
WS"
1610. 1615
/note="putative"
1629
/note="putative"
polyA_signal
1610. 1615
polyA_site
1629
/note="putative"
ORIGIN
Alignment Scores:
Pred. No.: 3,34e-126 Length: 1629
Score: 1227.50 Matches: 224
Percent Similarity: 76.32% Conservative: 79
Best Local Similarity: 56.42% Mismatches: 89
Query Match: 60.08% Indels: 5
DB: 3 Gaps: 2
US-10-039-183a-4 (1-399) x AK084724 (1-1629)
QY 2 AAlatggtgltuyspheasnaargThrlrpsprohisvalaenilegIyThrlleGIyhisval 21
DB 213 GCCAAGAAGACTTATGTAAGGACCAAGCCCATGTAATGGGTAACCATCGGACATGTG 272
QY 22 AsphlsglylvtrThrlleuSerAlaAlaleSerAlaValleuSerleuIyglIeu 41
DB 273 GACCAATGCAAGACCAACCTGACCTGACACATCAAGAAATTCATGACCGAGGAGTGTGA 332
QY 42 AAlagIuMetIySaSPtYrAspAsnIleAspAsnIleAspAsnIleAspAsnIleAspAsnIle 61
DB 333 GCTAAGTTCACAAAGACTAGAGGAGATGACATGCCCCAGAGAGAGAGAGAGAGAGAGAGAG 392
QY 62 ThrIleAlaThrSerHisIleGlyIyTrgluThrgluAenArghIstYrAlaHisIyAlaSp 81
DB 393 ACCATCATTCAGACGCCCATGTGAGATAGCACTGCTGCCGACACTATGCCACAGAC 452
QY 82 CysProGlyHisIaIaAspTYrValIyAsnMetIleThrgIyAlaAlaGImetCaspGly 101

Db 453 TGTCTGTGTCATGAGATTATGTATGAATATGATCATCAGGACCTGCCCCCTGGATGAC 512
 Oy 102 AAlaileuValValSerAlaAlaAspGlyProMetProGlnThrArgGlnHisIleu 121
 Db 513 TGTATCTGGTGGTGGACCTTAATGATGCCCCATGCCCCAAGACCAAGACCTTACG 572
 Oy 122 LeuSerArgGlnValGlyValProHisIleValPheLeuAsnLysGlnAspMetVal 141
 Db 573 CTAGCTAAACAGATTGGGGTATGACATGTTGTGTGTAGTGAACAAGACCAACGCTGTC 632
 Oy 142 AspArgGlnGluLeuGluLeuValGluMetGluValArgGluLeuSerAlaTyr 161
 Db 633 CAGGACTCAGAGATGTGAGTGTAGTCGAGTCGAGATCCGGAGCTGCTCAACCGATG 692
 Oy 162 GluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 181
 Db 693 GGCTATTAAGAGAGAACTCCAGCATCGTAGGCTCGGCTCTGTGTCCTTGAAGCA 752
 Oy 182 AAlaValAlaGlyAsnValGlyLutTropGlyGluLys--ValLeuLysLeuMetAlaGlu 200
 Db 753 CGTGACCT-----GAGCTAGGCGCGAAGTCACTGACAGAGCTCTGATGCT 800
 Oy 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
 Db 801 GTGACACCTCATCCAGTGCACCCGAGACCTGACCAAGCCCTTCTGCTCTGTA 860
 Oy 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgGlyLeuArgGly 240
 Db 861 GAGTCAGTCTACTTATTCCTGCGGGGACAGTGTGTACAGTACATTAAGCGTGGC 920
 Oy 241 ValValValValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThr 260
 Db 921 ATTTGAGAAAGAGAGATGATGATGATGCTGCGACATTAACAAGACATCCGCACTG 980
 Oy 261 ValThrGlyValGluMetPheArgGlyGluLeuGlyGlyGlyAlaGlyAspVal 280
 Db 981 GTGACAGCATTTGATGATGTTCCACAAGACCTGAGAGGCTGAGGACGAGGATTAACCTG 1040
 Oy 281 GlyValLeuLeuArgGlyThrIleValGluGluValGluArgGlyMetValLeuLys 300
 Db 1041 GGTGCTCTGCTCGAGGCTTAAAGCGCGCAAGATTGAGCGCTGTGCTCATGCTCAAG 1100
 Oy 301 ProGlySerIleThrProHisLysValPheGluGlyGluIleThrValLeuSerLysGlu 320
 Db 1101 CCAAGCTCATCAACCCCAACCAAGAGTGGAGCCGAGTTTATCTCTGCAAGAG 1160
 Oy 321 GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheThrValArgThr 340
 Db 1161 GAAAGAGCGCCCAACAACCTTGTATCTCATGCCCCGTCATGTTCCCGTAC 1220
 Oy 341 ThrAspValThrGlySerIleThrIleProGluGlyValGluMetValMetProGlyAsp 360
 Db 1221 TGGGACATGCGCTGAGTCTTCTTCCCTCCAGGAGAACTTGCCCTGCTGAGAG 1280
 Oy 361 AsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrIleAsp 380
 Db 1281 GACTTGAAGCTTATGCTATCTTGGCGGACCCCATGATTTTGAAGAAAGCCCAAGCTTTC 1340
 Oy 381 AlaIleArgGluGlyArgThrValAlaGlyValValSerAsnIle 397
 Db 1341 ACCTTGAAGGATGCAACAAGACATTTGGCACTGGCTTGTACGATGTA 1391
 RESULT 8
 CRS93603 1494 bp mRNA linear HTC 21-JUL-2004
 LOCUS full-length cDNA clone CSODK011YC01 of HeLa cells Cot 25-normalized
 DEFINITION of Homo sapiens (human).
 ACCESSION CRS93603
 VERSION CRS93603.1 GI:50474410
 KEYWORDS HTC; cNSLT; cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1494)
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 REMARK Contact : Feng Liang Email : fliang@life.techn.com URL :
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600
 Faraday Avenue
 2 (bases 1 to 1494)
 REFERENCE
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with NotI and cloned
 into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 FEATURES
 source location/Qualifiers
 1..1494
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSODK011YC01"
 /tissue_type="HeLa cells Cot 25-normalized"
 /plasmid="pCMVSPORT_6"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,816-126 Length: 1494
 Score: 1226.50 Matches: 228
 Percent Similarity: 75.38% Conservative: 72
 Best Local Similarity: 57.29% Mismatches: 93
 Query Match: 60.03% Indels: 5
 DB: 3 Gaps: 2
 US-10-039-183a-4 (1-399) x CRS93603 (1-1494)
 Oy 2 AAlaValGluLysPheAsnArgThrIleProHisValAlaGlnIleGlyThrIleGlyHisVal 21
 Db 97 GCCAAGAAAGACTTATGTCGCGCAACAGCCACATGTGAATGTGGTATCCATTCGCGCATG 156
 Oy 22 AspHisGlyLysThrThrLeuSerAlaAlaIleSerAlaValLeuSerLysGlyLeu 41
 Db 157 GACCAAGGAGAACCAAGCTGACTGACGACCATCAAGAGATTCTAGGAGGAGGTGG 216
 Oy 42 AlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGlyIle 61
 Db 217 GCTAAGTTCAGAAATACAGAGAGATTCAGATGCCCCGAGAGACGAGCTGGGGTATC 276
 Oy 62 ThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgIleTyrAlaHisValAsp 81
 Db 277 ACCATCAATGCGGCTCATGTGGATGTAGCACTGCCGCCGCACTACGCCCAACAGAC 336
 Oy 82 CysProGluHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGluMetAspGly 101
 Db 337 TGCCCGGGTCATGACGATTATGTATGAATATGATACAGGACCTGCACCTCGACGGC 396
 Oy 102 AAlaileuValValSerAlaAlaAspGlyProMetProGlnThrArgGlnHisIleLeu 121
 Db 397 TGCATCTCGTGTGAGACAGCCCAATGACGCGCCCAAGCCCAAGACCACTTATTA 456
 Oy 122 LeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMetVal 141
 Db 457 CTGGCCACACAGATTGGGGTGGAGCATGTGTGTGTATGTGAACAAGGCTCAACCTGTC 516
 Oy 142 AspArgGlnGluLeuGluLeuValGluMetGluValArgGluLeuSerAlaTyr 161
 Db 517 CAGGACTTGAAGTGTGAATGTGTGAACCTGGAAGTCCGGAGCTGACCAAGTTT 576
 Oy 162 GluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGluGlu 181

```

Db      577 GCGTATTAAGGGGAGAGAGACCCAGTCATCGTAGGCTCTGCTCTGTCGCTTGAG--- 633
Qy      182 AAlaySalagIyAenValgIyUtrpGlyUlys---ValleuyleuMeAlagIu 200
Db      634 -----GGTGGGAGCCCTGAGTTAGGCTTGAAGTCTGTCAAGAGCTACTGAGTCT 684
Qy      201 ValaSpAlaYrIleProThrProgluAraSpThrgIuYrThrPheleuMeProVal 220
Db      665 GTGACACTTACATCCAGAGTCCCGCCGGAGACCTGAGAAACCTTCTGCTGCTG 744
Qy      221 GluAraPValPheSerIleAlaIyArGlyThrValValThrgIyArGIlleGluArGly 240
Db      745 GAGGGGCTGTACTCCGCTCCCTGCGCGGACCGGTGTGACAGTACACTAGAGGTGGC 804
Qy      241 ValValyValgIyAaSpGluValgIuIleValgIyIleArpProThrGlnuYrThrThr 260
Db      805 ATTTTAAAGAGAGGAGACAGAGTGTAGCTCTTGAAGCATGACAGAAACATCCGACTGTG 864
Qy      261 ValThrgIyValgIuMeArPheArGlySgluLeuGlyUlysGlyUlaIagIyAaSpAla 280
Db      865 GTGACAGGATTCGATGATGTTCCAAAGAGCCTGAGAGAGGCGGAGGCGAGATTAACCTC 924
Qy      281 GlyValleuLeuArGlyThrIySgluSgluValgIuArGlyMeValleuCySlys 300
Db      925 GGGGGCTGTGTCGAGGCTTGAGAGCGGAGGACTTGCGCGGCTGCTCATGTCAG 984
Qy      301 ProgluSerIleThrProHislySlySpPhegluGlyUleYrValleuSerIySglu 320
Db      985 CCAGGTCATTAACCCCAAGAGAGGTGGAGGCCCGGATTACATCTCCAGCAAGAG 1044
Qy      321 GluIyGlyArGlyThrProPhePheThrAenYrArpProgluPheYrValArGThr 340
Db      1045 GAAGGTGGCCGCCCAAGAGCCTTGTGTCCTTCATGCTCATGATGTCCTCCGACT 1104
Qy      341 ThrAraPValThrgIySerIleThrIleuProgluGlyValgIuMeValleuProgluYr 360
Db      1105 TGGGACATGGCTCGATTAATCTGTCGCCCAAGAGAGACTTGCCATGCGCGGAG 1164
Qy      361 AenValySlyIleThrValgIuLeuIleSerProValAlaIleuGlyUleGlyThrIySph 380
Db      1165 GACCTGAAGTTACACTTAATCTTGCGGAGACCAATGATCTTGAAGAAAGCCAGCTTTC 1224
Qy      381 AlaIleArGlyUlyGlyArGThrValgIyAlaIyAlaIyValSerAenIlele 398
Db      1225 ACCCTGCGAGATGACACCGGACTATGCGACCGGTCACTACCAACACCGCTG 1278

RESULT 9
LOCUS   CR621980               1505 bp      mRNA       linear   HTC 21-JUL-2004
DEFINITION   Full-length cDNA clone CS0DJ015YL11 of T cells (Jurkat cell line)
ACCESSION   CR621980
VERSION     CR621980.1  GI:50502787
KEYWORDS    HTC; cNSLT cDNA.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 1505)
AUTHORS    Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
CONTACT     Contact : Feng Liang Email : fliang@lifetech.com URL :
REMARK     http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue
2 (bases 1 to 1505)
REFERENCE   Direct Submision
AUTHORS    Genoscope.
TITLE      Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL    BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned

```

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

source

1. 1505

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DJ015YL11"

/tissue_type="T cells (jurkat cell line) Cot

10-normalized"

/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.:	3,856-126	Length:	1505
Score:	1226.50	Matches:	228
Percent Similarity:	75.38%	Conservative:	72
Best Local Similarity:	57.29%	Mismatches:	93
Query Match:	60.03%	Indels:	5
DB:	3	Gaps:	2

US-10-039-183a-4 (1-399) x CR621980 (1-1505)

```

Qy      2 AAlaySgluYrPheAaArGThrIyPProHisValAenIleGlyThrIleGlyAenVal 21
Db      132 GCCAAGAAAGTACTAGTCCGCGAGACAGCCACATGATGATGAGGTACCATGCGCATGTG 191
Qy      22 AaPhIleGlyYrThrThrIleuSerAlaAlaIleSerAlaValleuSerIleuYr 41
Db      192 GACCAAGGAGACCAAGCTTACCTGACACCATCAAGAAATTCAGCTGAGAGAGTGG 251
Qy      42 AlaIuMeIleAaPyrYrAaSpAenIleAaSpAlaPProgluGlyUlaArGlyIle 61
Db      252 GCTAAGTTCAGAAAGTACAGAGATTAACAATGCCCCGAGAGACAGCTCGGGATTC 311
Qy      62 ThrIleAlarSerHisIleGlyUlyrGlyUthrgIuAaArGThrIyAlaIleValAaP 81
Db      312 ACCATCAATGGGCTCATGTGATGATGACATGCGCGCCGCTACAGCCACACAGAG 371
Qy      82 CySPProgluYrHisAaPyrYrValIyAaAenMetIleThrgIyAlaIleGluMeArSpGly 101
Db      372 TGCCGGGTCTATGAGATTAATGATGATGATGATGATGATGATGATGATGATGATG 431
Qy      102 AlaIleuValAlaIleSerAlaAlaAaSpGlyProMetProgluThrArGlyUhiIleleu 121
Db      432 TGCACTCTGCTGTAGACAGCCAAATGACGCGCCCATGCCCCAGACCCGAGACCTTATTA 491
Qy      122 LeuSerArGluValgIyValProHisIleValAlaPheleuAaenYrSgluAaPMeVal 141
Db      492 CTGGCAGACAGATTTGGGGGTGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 551
Qy      142 AaPhePgluIleuLeuGlyUleuValgIuMeGlyUlaArGlyUleuLeuSerAlaYr 161
Db      552 CAGACTCTGAGAGGTGTGACCTGTGTGACCTGTGAGATCCCGAGACTGTCTACCAAGTTT 611
Qy      162 GluPheProgluYrAaPArThrProIleValAlaIySerAlaIleuArGAlaIleuGlyU 181
Db      612 GCGTATTAAGGAGAGAGACCCAGCTATCGTAGGCTGTGCTCTGTGCGCTTGAG--- 668
Qy      182 AAlaySalagIyAenValgIyUtrpGlyUlys---ValleuyleuMeAlagIu 200
Db      669 -----GGTGGGAGCCCTGAGTTAGGCTTGAAGTCTGTGCAAGAACTACTGATGCT 719
Qy      201 ValaSpAlaYrIleProThrProgluAraSpThrgIuYrThrPheleuMeProVal 220
Db      720 GTGACACTTACATCCAGAGTCCCGCCGGAGACCTGAGAAAGCCTTCTCTGCTGCTGTG 779
Qy      221 GluAraPValPheSerIleAlaIyArGlyThrValValThrgIyArGIlleGluArGly 240
Db      780 GAGGGGCTGTACTCCGCTCCCTGCGCGGACCGGTGTGACAGTACACTAGAGGTGGC 839
Qy      241 ValValyValgIyAaSpGluValgIuIleValgIyIleArpProThrGlnuYrThrThr 260

```

```

Db      840 ATTTAAAGAGGAGAGAGAGTGTGCTCTAGGACATAGCAAGAACATCCGCACTGCG 899
      :::::|||||:::|||||:::|||||:::|||||
Qy      261 ValThrGlyValGluMetPheArgGlyGluGluValGlyValGlyValGlyValGlyVal 280
      :::::|||||:::|||||:::|||||:::|||||
Db      900 GTGACAGGATTTGAGATGTTCCACAGACCTGAGAGGCGGAGGCCGAGATTAACCTTC 959
      :::::|||||:::|||||:::|||||:::|||||
Qy      281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuGlyLys 300
      :::::|||||:::|||||:::|||||:::|||||
Db      960 GGGGCGCTGTGCGAGCTTGAAGCGGAGAGCTTGCGGCGGCTGCTGATGCTCAAG 1019
      :::::|||||:::|||||:::|||||:::|||||
Qy      301 ProGlySerIleThrProHisLysLysPheGluGluGluIleThrValLeuSerLysGlu 320
      :::::|||||:::|||||:::|||||:::|||||
Db      1020 CCAGGTTTCATCAAGCCCAAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1079
      :::::|||||:::|||||:::|||||:::|||||
Qy      321 GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGluPheTyrValArgThr 340
      :::::|||||:::|||||:::|||||:::|||||
Db      1080 GAAAGGTGGCCGACAGAGCCCTTGTGTCCCATTCATGCTGTCTGCTCCCTGACT 1139
      :::::|||||:::|||||:::|||||:::|||||
Qy      341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValLeuProGlyAsp 360
      :::::|||||:::|||||:::|||||:::|||||
Db      1140 TGGGACATGGCTGTGCGATTAATCTGCCCCAGAGAGAGAGCTTGCATGCGGAGAG 1199
      :::::|||||:::|||||:::|||||:::|||||
Qy      361 AsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrLysPhe 380
      :::::|||||:::|||||:::|||||:::|||||
Db      1200 GACCGAAGTTCAACCTATCTTGGCGGACGCAATGATCTTAGAAGAGCCAGCGCTTTC 1259
      :::::|||||:::|||||:::|||||:::|||||
Qy      381 AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValLysSerAsnIleIle 398
      :::::|||||:::|||||:::|||||:::|||||
Db      1260 ACCCTGCGAGATGAGCAACCGGACTATTGGCACCGGCTGATGACCAACACGCTG 1313
      :::::|||||:::|||||:::|||||:::|||||

```

RESULT 10

LOCUS CR608035 1534 bp mRNA linear HTC 21-UTL-2004
DEFINITION full-length cDNA clone CS0D1007YB10 of Placenta Cot 25-normalized
of Homo sapiens (human).

ACCESSION CR608035
VERSION CR608035.1 GI:50488842
KEYWORDS HTC; cNSUT_cDNA
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1534)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradey Avenue
Foster City, CA 94024
2 (bases 1 to 1534)
Genoscope.

REFERENCE 2
AUTHORS Direct Submission
TITLE Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail: sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source location/Qualifiers

1..1534
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1007YB10"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.: 3,966-126 Length: 1534

Score: 1226.50 Matches: 228
Percent Similarity: 75.38% Conservative: 72
Best Local Similarity: 57.29% Mismatches: 93
Query Match: 60.03% Indels: 5
DB: 3 Gaps: 2

US-10-039-183a-4 (1-399) x CR608035 (1-1534)

```

Qy      2 AlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHisVal 21
      :::::|||||:::|||||:::|||||:::|||||
Db      126 GCCAAGAGACTTACGTGGCGACAGACCAATGTGAATGTGAATGTGAATGTGAATGTGAATGTG 185
      :::::|||||:::|||||:::|||||:::|||||
Qy      22 AspHisGlyLysThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLysGlyLeu 41
      :::::|||||:::|||||:::|||||:::|||||
Db      186 GACCAAGGAGAGACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
      :::::|||||:::|||||:::|||||:::|||||
Qy      42 AlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGlyIle 61
      :::::|||||:::|||||:::|||||:::|||||
Db      246 GCTAAGTTCAGAGATGACAGAGAGATTGACATGCCCGGAGAGAGAGAGAGAGAGAGAGAGAG 305
      :::::|||||:::|||||:::|||||:::|||||
Qy      62 ThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisValAsp 81
      :::::|||||:::|||||:::|||||:::|||||
Db      306 ACCATCATATGCGCTCATGTGATGATAGACACTGCGCCCGCCACTAGCCCAACAGAC 365
      :::::|||||:::|||||:::|||||:::|||||
Qy      82 CysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGluMetAspGly 101
      :::::|||||:::|||||:::|||||:::|||||
Db      366 TCCCGGGCTATGCAATTAATTTATTAAGATATGATATCAAGGACATGACACCTCCAGAGGC 425
      :::::|||||:::|||||:::|||||:::|||||
Qy      102 AlaIleLeuValLysSerAlaAlaAspGlyProMetProGluThrArgGluHisIleLeu 121
      :::::|||||:::|||||:::|||||:::|||||
Db      426 TGCATCCGTGTGTGAGACGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 485
      :::::|||||:::|||||:::|||||:::|||||
Qy      122 LeuSerArgGluValGlyValProHisIleValAlaPheLeuAsnLysGluAspMetVal 141
      :::::|||||:::|||||:::|||||:::|||||
Db      486 CTGGCGACAGCATGATGGGTGAGCATGTGGGTGATGATGATGATGATGATGATGATGATGATG 545
      :::::|||||:::|||||:::|||||:::|||||
Qy      142 AspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAlaTyr 161
      :::::|||||:::|||||:::|||||:::|||||
Db      546 CAGGACTCTGAGATGTGAGACTGTGAGACTGTGAGACTGTGAGACTGTGAGACTGTGAGACTGT 605
      :::::|||||:::|||||:::|||||:::|||||
Qy      162 GluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGluGlu 181
      :::::|||||:::|||||:::|||||:::|||||
Db      606 GCGTATTAAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
      :::::|||||:::|||||:::|||||:::|||||
Qy      182 AlaLysAlaGlyAsnValGlyGluTyrGlyGlyLysValLeuLeuMetAlaGlu 200
      :::::|||||:::|||||:::|||||:::|||||
Db      663 -----GGTCGGGAGCCCTGAGTGAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 713
      :::::|||||:::|||||:::|||||:::|||||
Qy      201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
      :::::|||||:::|||||:::|||||:::|||||
Db      714 GTGACACTTACATCCAGTGCCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 773
      :::::|||||:::|||||:::|||||:::|||||
Qy      221 GluAspValPheSerIleAlaGlyArgGlyThrValAlaThrGlyArgGlyIleGluArgGly 240
      :::::|||||:::|||||:::|||||:::|||||
Db      774 GAGGCGGTGTACTCCGTCCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 833
      :::::|||||:::|||||:::|||||:::|||||
Qy      241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGluLysThrThr 260
      :::::|||||:::|||||:::|||||:::|||||
Db      834 ATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 893
      :::::|||||:::|||||:::|||||:::|||||
Qy      261 ValThrGlyValGluMetPheArgLysGluGluGluValGlyValGlyValGlyValGlyVal 280
      :::::|||||:::|||||:::|||||:::|||||
Db      894 GTGACAGGATTTGAGATGTTCCACAGACCTGAGAGGCGGAGGCCGAGATTAACCTTC 953
      :::::|||||:::|||||:::|||||:::|||||
Qy      281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuGlyLys 300
      :::::|||||:::|||||:::|||||:::|||||
Db      954 GGGGCGCTGTGCGAGCTTGAAGCGGAGAGCTTGCGGCGGCTGCTGATGCTCAAG 1013
      :::::|||||:::|||||:::|||||:::|||||
Qy      301 ProGlySerIleThrProHisLysLysPheGluGluGluIleThrValLeuSerLysGlu 320
      :::::|||||:::|||||:::|||||:::|||||
Db      1014 CCAGGTTTCATCAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1073
      :::::|||||:::|||||:::|||||:::|||||
Qy      321 GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGluPheTyrValArgThr 340
      :::::|||||:::|||||:::|||||:::|||||

```

Db 1074 GAAGGTGGCCGCCACAGCCCTTTGTGTCACCTTCATGCTGATGTTCTCCCTGACT 1133
Qy 341 ThrAspVal1ThrGlySer1LeuPheProGluGlyValGluMetValMetProGlyAsp 360
Db 1134 TGGGACATGGCCCTGCTGGATTTCTCTGCCCCCAGAGAGAGACCTTGCCATGCCCCGGAG 1193
Qy 361 AsnVal1ySileThrValGluLeuLeuSerProValAlaLeuGluLeuGlyThrIleAsp 380
Db 1194 GACCTGAAGTTCACCTTAATCTTGGCGGACGCAATGATCTTTAGAGAAAGGCCAGCGTTTC 1253
Qy 391 AlaIleArgGluGlyValArgThrValGlyAlaGlyValValSerAsnIleIle 398
Db 1254 ACCCTGCGAGATGACACCGGACTATTGGCACCGGTCTGATCACCACACGCGTG 1307

RESULT 11
LOCUS AY105509 1193 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO100031 mRNA sequence.
ACCESSION AY105509
VERSION AY105509.1 GI:21208587
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 1193)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whiteitt,M.S., Arthur,L.W., Hainey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
Unpublished (2002)
2 (bases 1 to 1193)
Coe,E.H.
Direct Submission
Submitted (25-Apr-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones, these are publicly available from ZMBD and may be found by BLAST searching at MSU, maizeMap.org; ZMBD, www.zmbd.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize CDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZMBD: www.zmbd.iastate.edu.

FEATURES
source location/Qualifiers
1..1193
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

ORIGIN
Alignment Scores:
Pred. No.: 1,48e-111 Length: 1193
Score: 1095.00 Matches: 215
Percent Similarity: 80.50% Conservative: 41
Best Local Similarity: 67.61% Mismatches: 50
Query Match: 53.60% Indels: 12
Gaps: 3
US-10-039-183a-4 (1-399) x AY105509 (1-1193)
Qy 93 lIeThrgIyAlaAlaGluMetAspGlyAlaIleLeuValValSerAlaAlaAspGlyPro 112

Db 3 ATCACCGCGCGCTGCGCANATGACGCGCCATCTCTGCTGCTGCGCGCGGACGTCCTC 62
Qy MetProGlnThrArgGluHisIleLeuLeuSerArgGluValGlyValProHisIleVal 132
Db 63 ATGCCGACACCAAGACACACATCTCTCCGCAAGAGAGTGGGTTCCCAAGATCGTT 122
Qy 133 ValPheLeuAsn1ySgInAspMetValAspAspGlnGluLeuLeuValGluMet 152
Db 123 GTCTTCCTCAACAGAAAGACATGCTGACGACGAGAGCTGCTGACCTGTCGAGCTG 182
Qy 153 GluValArgGluLeuLeuSerAlaTyrGluPheProGlyAspAspThrProIleValAla 172
Db 183 GAAGTCCCGGAGCTGCTCAGCACTATGATGACGCGGACGAGTGCATATGTCCT 242
Qy 173 GlySerAlaLeuAspGlyAlaLeuGluGluAlaGlyAsnVal----- 187
Db 243 GACTCCGCTCAAGGCGCTC--GAGGCTCTCATGGGCAACCTTACCTTGAAGCGCGC 299
Qy 188 --GlyGluTyrPrgGlyGluValLeuLeuLeuMetAlaGluValAspAlaTyrIlePro 206
Db 300 GACGATGATGGGTGACATGATCTTCAAGCTGGTGTATCTGTGATTCCTACATTC 359
Qy 207 ThrProGluArgAspThrGluLeuThrPheLeuMetProValGluAspValPheSerIle 226
Db 360 GTGCCACAGCGCGACCGACCTCCCTCTTACTGCTGTGAAGATGTCCTTCCATC 419
Qy 227 AlaGlyArgGlyThrValThrGlyValArgIleGluArgGlyValValValValGlyAsp 246
Db 420 ACCGTCGTGTGACGTTGCGCATCTGGCCGCGATCGAGCTGGACCGGTCAAGATGCTATC 479
Qy 247 GluValGluIleValGlyIleArgProThrGlnIleThrThrValThrGlyValGluMet 266
Db 480 AAGTCGATATCGTTGGATCCGGGACACCGGAACTGCACCGTACCTGGGCTTGAGATG 539
Qy 267 PheArgGlyGluLeuGluLeuGlyGluAlaGlyAspAsnValGlyValLeuLeuArgGly 286
Db 540 TTCGAGAGACCATGACGATGATGCTGAGCAATGTCGAGCTGCTGCTGCTGCTGCTG 599
Qy 287 ThrIleSlyGluGluValGluValGluMetValLeuGlySlyProGlySerIleThrPro 306
Db 600 ATGCAGAAAGATGATGAGAGAGGAGTGGTGGCGAAGCCGGCTCTATACACCG 659
Qy 307 HisIleSlyPheGluGluIleThrValLeuSerIleSlyGluGluGlyIleArgHisThr 326
Db 660 CACACCAAGTTGAGCGCTGTGTATGCTTAAGAGAGAGCGGTGGCGGACATCA 719
Qy 327 ProPhePheThrAsnTyrArgProGlnPheThrValArgThrThrAspValThrGlySer 346
Db 720 CTTTCTTCCCTGTTACCGCCGCACTTCAATGAGACACCGATGTGACAGCGGAT 779
Qy 347 lIeThrLeu-----ProGluGlyValGluMetValMetProGlyAspAsn 361
Db 780 GTGACTGTGATTATGATGACAAAGATGAGAGCGGAAAGTGTGATCTGGTGACCT 839
Qy 362 ValIleSileThrValGluLeuLeuSerProValAlaLeuGluLeuGlyThrIlePheAla 381
Db 840 ATCAAAGATGTTGTTCACTATCAGCTGTGCTTGTGTGAGCAGGATATGAGTTTCT 899
Qy 392 lIeArgGluGlyValArgThrValGlyValAlaGlyValValSerAsnIleIleGlu 399
Db 900 ATCCGTGAGGTGTGAAGACTGTGCTGCTGTCATCAACAATAATCATTTAG 953

RESULT 12
CK215616/c 1119 bp mRNA linear EST 09-DEC-2003
LOCUS FGAS027583 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum
DEFINITION aestivum cDNA, mRNA sequence.
ACCESSION CK215616
VERSION CK215616.1 GI:39621720
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

COMMENT

Contact: Rod A. Wing
Arizona Genomics Institute
The University of Arizona
Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
Tel: 520 626 9595
Tel: 520 626 9595
Fax: 520 621 1259
Email: <http://genome.arizona.edu>
Plate: 48 row: H column: 17.

FEATURES

1.899

ORIGIN

Alignment Scores:

Pred. No.:	3.32e-108	Length:	892
Score:	1063.50	Matches:	199
Percent Similarity:	82.81%	Conservative:	37
Best Local Similarity:	69.82%	Mismatches:	44
Query Match:	52.06%	Indels:	5
DB:	7	Gaps:	1

US-10-039-183A-4 (1-399) x CO083417 (1-899)

[illegible]

QY	263	GIYVALIGIMETPhaXIGLysGIueGIueGIuLYsGIuVALIGIYAspAsnVALIGIYVAL	282
Db	620	GGTGGGAAAGTGTTCAAAGACATTAGATAGACGCCATGGCTGGTATATATGGGCTTG	679
QY	283	LeuLeuAARGIYThrLYbysGIuGIuVALIGIuARGIYmeLValIeuCYbLYsPProGIY	302
Db	680	TTACTTAAGGCTGTCAAAAGGCCGATATTACAGAGGGATGTTTGGCTAAACAGGG	739
QY	303	SerIleThrProHISLYbysPheGIuGIuIeTYrVALIeuSerYsGIuGIuGIY	322
Db	740	ACAACTACTCCCCAATACCAGATTCAGCCGCGCATTTGTATGTGTGTAAGAGAAAGAGGGT	799
QY	323	GIYARGHISThrProPhePheThraSNYrARGProGIuPheTYrVALrGThrThrAsP	342
Db	800	GGCAGACATTCCTCCCTTCGCGAGTATATAGAGCCCTCAAGTTCATCAGAGACCACTGAT	859
QY	343	ValThrNGIYSerIle	347
Db	860	GTCACTGGAAGGGTG	874

RESULT 15

LOCUS	CV291218	879 bp	mRNA	linear	EST 23-SEP-2004
DEFINITION	aof01-11m01-g03 Aof01 Asparagus officinalis cDNA clone				
ACCESSION	aof01-11m01-g03 5', mRNA sequence.				
VERSION	CV291218				
KEYWORDS	CV291218.1	GI:56577294			
SOURCE	EST.				
ORGANISM	Asparagus officinalis (garden asparagus)				
	Asparagus officinalis				

FEATURES

SOURCE

ORIGIN

Alignment Scores:

Pred. No.: 1.51e-107 Length: 879
 Score: 1057.50 Matches: 202
 Percent Similarity: 82.53% Conservative: 39
 Best Local Similarity: 69.18% Mismatches: 44
 Query Match: 51.76% Indels: 7
 DB: 7 Gaps: 2

US-10-039-183a-4 (1-399) x CV291218 (1-879)

```

OY 50 ILeaPAnA1AProGluGluPheGluValArgGlyIleThrIleAlaThrSerHisIleGlu 69
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7 ATCGACGCCGCGCCCGAGAGCGCGCCCGTGGCATCACCATCAACACCGCCACCGTCGAG 66
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 70 TyrGluThrGluAsnArgHisIleTyrAlaHisValAspCysProGlyHisAlaAspTyrVal 89
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67 TACGAGACCGAAGATCGCCATTACGCCACGTCGACTGCCGCCGACGACGATTAATGTC 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 90 LysAsnMetIleThrGlyAlaAlaGlnMetAspGlyAlaIleLeuValAlaSerAlaAla 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 127 AAGAACATGATCACCGCGCGCTCCAGATGACGCGGTCATTCTGCTGTCGCGGCT 186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 110 AspGlyProMetProGluThrArgGluHisIleLeuLeuSerArgGluValGlyValPro 129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 187 GACGGGCCAATGCCGAGACCAAGAGCATATCTGCTTGCGAAGCAGGTCGGGGTGCTT 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 130 HisIleValAlaPheLeuAsnIleGluAspMetValAspAspGlnGluLeuLeuGluLeu 149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 247 AACATGTCGTGCTTCTTGAATAAGCAGACGACGTCGATGATGAGGAGCTTCTGGAAATTG 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 150 ValGluMetGluValArgGluLeuLeuSerAlaIleTyrGluPheProGlyAspAspThrPro 169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 307 GTGGAGCTCGAGGTCAAGGAGCTGCTTACTCTCAGAATTCGCCGCGATGATGTCCT 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 170 IleValAlaGlySerAlaLeuArgAlaLeuGlu-----GluAlaLys 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 367 ATTGTATCGGGCTCTGCACTGTAGCTTTAGAACCATGATGAAGATCCCGCGTTGAGG 426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 184 AlaGlyAsnValGlyIleTyrGlyIleValLeuLysLeuMetAlaGluValAlaAspAla 203
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 427 CGGGCGCAG---GATCAGTGGGTGACAAAGATCTACGATCTAATGACAAAGTGACAGC 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 204 TyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProValGluAspVal 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 484 TACATTCCGATTCTCCAGAGCGACGACCGATCTCCCTTCTCTCTGCTAATGAGGATGTG 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 224 PheSerIleAlaGlyArgGlyThrValAlaThrGlyArgIleGluArgGlyValValLys 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 544 TTTTCATCAACGGGTGCGGACGCGTGGCCACTGGCCGTTGGAGAGAGGACCATTCAG 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 244 ValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThrValThrGly 263
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 604 GTCGGAGAAACGTGACACTTGTGCTCTCAAGGAGACACGAGACACCATGTGACCGGC 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 264 ValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnValGlyValLeu 283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 664 GTCGAGATGTTCCAAAGATTCGTGACGACGATATGCGCAGGACCAATGTTGATTGCTG 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 284 LeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLysProGlySer 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 724 CTTGCTGGGATTCAGAGAGCGGATATTCAGAGGAGATGTGCTINGGCTAGCCGGCCACC 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 304 IleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGluGluGlyGly 323
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 784 ATCACTCCCAACACCAAGATTCCTGCGATGTGTTATGTGCTTAAGAGAGAGAGGGGGGT 843
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 324 ArgHisThrProPhePheThrAsnTyrArgProGln 335
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 844 CGCCACTCGCATTCNTTCGTGGGTACAGACTCTCAG 879
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

Search completed: April 24, 2005, 11:20:20
 Job time : 4496.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: Apr11 24, 2005, 03:22:05 ; Search time 6029.59 Seconds
(without alignments)
3206.460 Million cell updates/sec

Title: US-10-039-183A-4
Perfect score: 2043
Sequence: 1 MAKEKFRNTKPHVNIQTIGH.....FAIREGRTVAGVSNITIE 399

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-O=/cgn2.1/USPTO/spool/US10039183/runat.22042005.122425.20111/app.query.fasta.1.1038
-DB=genembl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOEXT=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human0.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO MMAP -LARGESUBSTR -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
Genembl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pal:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_pro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2043	100.0	1448	6	BD082347
2	2043	100.0	10532	6	BD000626 Helicobac
3	2035	99.6	2825	6	BD009865 Proteins
4	2001	97.9	15292	1	AE001541 Helicobac

c	5	1900	93.0	308929	1	AE017145	AE017145 Helicobac
c	6	1862	91.1	346792	1	BX571658	BX571658 Wolinella
c	7	1831.5	89.6	1203	1	W51740TUF	X76872 W. ruscinioge
c	8	1772	86.7	1865	1	CUT17167	Y17167 Campylobac
c	9	1772	86.7	308601	1	C111168X2	AL139075 Campyloba
c	10	1649.5	80.7	1185	12	AY305395	AY305395 Synthetic
c	11	1641.5	80.3	1185	12	AY305397	AY305397 Synthetic
c	12	1622.5	79.4	1185	12	AY305396	AY305396 Synthetic
c	13	1611.5	78.9	204050	1	AL646073	AL646073 Ralstonia
c	14	1609.5	78.8	7934	1	AE005800	AE005800 Caulobac
c	15	1609.5	78.8	10212	1	AE005984	AE005984 Caulobac
c	16	1609.5	78.8	10494	1	AE002372	AE002372 Neisseria
c	17	1609.5	78.8	349980	6	AX043922	AX043922 Sequence
c	18	1608	78.7	1185	6	CO817257	CO817257 Sequence
c	19	1608	78.7	326434	1	AE016947	AE016947 Enterococ
c	20	1606.5	78.6	10978	1	AE002371	AE002371 Neisseria
c	21	1606.5	78.6	340806	1	NMA122491	AL162732 Neisseria
c	22	1605.5	78.6	10293	1	AE010465	AE010465 Fusobacte
c	23	1604.5	78.5	1191	1	TCHESTUP	X76871 T. cupritus
c	24	1604.5	78.5	110000	1	AE017180	Continuation (32 o
c	25	1601	78.4	1302	1	BSYRGNE	Continuation (32 o
c	26	1600	78.3	1185	6	CO817259	Continuation (32 o
c	27	1594	78.0	301727	1	AE016960	Continuation (32 o
c	28	1593.5	78.0	301457	1	AE016924	Continuation (39 o
c	29	1592.5	77.9	110000	1	BX571665	Continuation (28 o
c	30	1592.5	77.9	110000	1	CP000010	Continuation (12 o
c	31	1592	77.9	305050	1	AL596173	Continuation (12 o
c	32	1592	77.9	311208	6	AX417049	Continuation (12 o
c	33	1592	77.9	349980	6	AX417036	Continuation (12 o
c	34	1590.5	77.9	286850	1	BX321857	Continuation (12 o
c	35	1590.5	77.9	313050	1	BX321857	Continuation (12 o
c	36	1587.5	77.7	110000	1	AE017282	Continuation (12 o
c	37	1586.5	77.7	19552	1	AE013173	Continuation (12 o
c	38	1586	77.6	225228	1	AL591864	Continuation (12 o
c	39	1586	77.6	244528	6	AX641673	Continuation (12 o
c	40	1586	77.6	273869	1	AE017331	Continuation (12 o
c	41	1586	77.6	349980	6	AX641672	Continuation (12 o
c	42	1585.5	77.6	1191	1	TCU78306	Continuation (12 o
c	43	1584.5	77.6	5203	1	TCU78306	Continuation (12 o
c	44	1584.5	77.6	12161	1	AE013172	Continuation (12 o
c	45	1584.5	77.6	110000	1	AE017333	Continuation (12 o

ALIGNMENTS

RESULT 1	BD082347	1448 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	BD082347	76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and			
DEFINITION	BD082347	corresponding polynucleotide molecules.			
ACCESSION	BD082347	GI:22627957			
VERSION	BD082347.1	GI:22627957			
KEYWORDS	JP 2001523954-A/54.				
SOURCE	Mastradenovirus				
ORGANISM	Mastradenovirus				
REFERENCE	1 (bases 1 to 1448)				
AUTHORS	Kleanthous, H., Lisso, L., Tomb, J.F., Miller, C. and Garawi, A.A.				
TITLE	corresponding polynucleotide molecules				
JOURNAL	Parent: JP 2001523954-A 54 27-NOV-2001;				
COMMENT	MERIEUX ORAYAK SOCIETE EN NOM COLLECTIF PASSEUR MERIEUX SERIUS ET VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA, HUMAN GENOME SCIENCES INC				
PN	JP 2001523954-A/54				
PD	27-NOV-2001				
PF	31-MAR-1998 JP 1998541962				
PI	HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI MILLER,				
PI	ANAL. AL GARAWI				
PC	ADIN43/04, ADIN59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02, A61K39/40,				
PC	GOIN33/554, GOIN33/569				

CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
 source
 Location/Qualifiers
 1..1448
 /organism="Mastadenovirus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10509"
ORIGIN
Alignment Scores:
Pred. No.: 1.1e-127 Length: 1448
Score: 2043.00 Matches: 399
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
 Gaps: 0
US-10-039-183a-4 (1-399) x BD082347 (1-1448)
QY 1 MetAlaLVsGluLVsPheAspAArgThrIleValAsnIleGlyThrIleGlyHis 20
Db 118 ATGGCAAAAGAAAGTTTAAACAGAACTAAAGCCGATGTAATTAATGAAACCAATGGGCAAT 177
QY 21 ValAspHisGlyLVsGlyThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLVsGly 40
Db 178 GTAGACCAATGCTAAACAGCACTTGAGTGCAGCGATTTGCGCGTCTTTCTTGAAAGGT 237
QY 41 LeuAlaGluMetLVsAspTyrAspAsnIleAspAsnAlaProGluGluLVsGluLVsGly 60
Db 238 CTTCGAGAAATGAAAGACTATATATATGATTAACGCCCTGAAAGAAAGAAAGAGG 297
QY 61 IleThrIleAlaThrSerHisIleGlyThrGluLVsAspArgHisIleValIleVal 80
Db 298 ATCACTATGCTACTCTTCACATGTAATGAGACTGAAACAGACACTATCGCATGTG 357
QY 81 AspCysProGlyHisAlaAspTyrValLVsAspMetIleThrGlyAlaAlaGluMetAsp 100
Db 358 GATTGCCAGAGACAGCGTGCATATGTAATAAACAATGATACCGGTGCGCGCAATGAC 417
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
Db 418 GGAGCGATTTGTTGTTCTGCTGAGTGTGCGCCATATCCCAAACTAGGAGCATATC 477
QY 121 LeuLeuSerArgGlyValGlyValProHisIleValIleValPheLeuAsnLVsGluAspMet 140
Db 478 TTATGTCTCGTCAAGTAGGCGTGCCTCACATCGTGTGTTCTTTAAACAAAGACATG 537
QY 141 ValAspAspGlnGluLeuLeuGluLVsGluMetGluValArgGluLeuLeuSerAla 160
Db 538 GTAGATGACCAAGAAATTTGTTGAACTTGTAGAAATGGAAGTCCGGAATTTGTAAGCCGG 597
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 598 TATGATTTTCCCTGGCGATACACTCTCTATCGTAGCGGCTTACGCTTTAAGACCTTTAGAA 657
QY 181 GluAlaLVsAlaGlyAsnValGlyLVsTyrPheGlyLVsValLeuLeuSerLeuIleGlu 200
Db 658 GAAGCAAAAGCGTGTAAATGTGGTAGATGGGAGTGAAGAAAGTCTTAAACTTATATGCTGAA 717
QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLVsThrPheLeuMetProVal 220
Db 718 GTGGATGCTTATATCCCTACTCCAGAAAGACACTGAAAGAAACCTTCTGTATGCGCGT 777
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluLVsGly 240
Db 778 GAAGATGTCTTCTATTGCGGGTAGAGGACTGTGTTACAGTAGATGAAGAGGCG 837
QY 241 ValValLVsValGlyAspGluValGluIleValGlyIleArgProThrGlnLVsThrThr 260
Db 838 GTGGTGAAGTAGGCGATGAAGTGAAGAACTGTGGTATACACTTAACAAAGAAAGCGACT 897
QY 261 ValThrGlyValGluMetPheArgLVsGluLeuGluLVsGlyAlaGlyAspAsnVal 280

Db 898 GTAAACCGGTGTAAGAAATGTTTAGAAAGAGTTGAAAGAGGAAAGCCGCGCATATATG 957
QY 281 GlyValIleLeuValAspGlyThrIleLVsGluGluValAlaGlyLVsMetValLeuLVs 300
Db 958 GCGGTGCTTTTGAAGAGAACTTAAAGAAAGAGGAAACCGGTATGTTTATGCAAA 1017
QY 301 ProGlySerIleThrProHisIleLVsPheGluGlyLVsIleValLeuSerLVsGlu 320
Db 1018 CCAAGTTCTATACCTCCCAACAGAAATTTAGGGAGAAATTTATGCTCTTTCTAAAGA 1077
QY 321 GluGlyLVsArgHisIleThrProPheThrAsnTyrArgProGlnPheTyrValArgThr 340
Db 1078 GAAGCGGGAGACACACTCCATTTCTTCAACAATTCACCCCGCAATTTATGTGCGACA 1137
QY 341 ThrAspValIleThrGlySerIleThrLeuProGluGlyValAlaGluMetValMetProGlyAsp 360
Db 1138 ACTGATGTGACTGCTCTATACCCCTCTCTGAAGCGTGAAGAAATGTTATGCTGGCGAT 1197
QY 361 AsnValLVsIleThrValGluLeuIleSerProValAlaLeuGluLVsGlyThrLVsPhe 380
Db 1198 AATGTGAATATCATCTGATGATGATTAAGCCCTGTGGCTTAGACTTGGAACTAAATTT 1257
QY 391 AlaIleArgGluGlyLVsArgThrValGlyAlaGlyValValSerAsnIleIleGlu 399
Db 1258 GCGATTCTGAAGGCGGTAGACCGTGTGCTGTGTTGAGACCAATATTTATGAA 1314
RESULT 2
AE000626/c 10532 bp DNA linear BCT 06-APR-1999
LOCUS Helicobacter pylori 26695 section 104 of 134 of the complete genome.
DEFINITION Helicobacter pylori 26695 section 104 of 134 of the complete genome.
ACCESSION AE000626 AE000511
VERSION AE000626.1 GI:2314360
XREFS
SOURCE Helicobacter pylori 26695
ORGANISM Helicobacter; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.
REFERENCE 1 (bases 1 to 10532)
Tomb,J.-F., White,O., Kurlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khaila,H.G., Glodek,A., McKenney,K., Fitzgerald,J.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Uterback,T.R., Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
The complete genome sequence of the gastric pathogen Helicobacter pylori
JOURNAL Nature 388 (6642), 539-547 (1997)
MEDLINE 97394467
PUBMED 9252185
REFERENCE 2 (bases 1 to 10532)
Tomb,J.-F., White,O., Kurlavage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khaila,H.G., Glodek,A., McKenney,K., Fitzgerald,J.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Uterback,T.R., Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M., Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
Direct Submission
JOURNAL Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 10532)
White,O.
TITLE Direct Submission
AUTHORS Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
JOURNAL location/Qualifiers
FEATURES
 source
 1..10532

gene /organism="Helicobacter pylori 26695"
/mol_type="genomic DNA"
/strain="26695"
/db_xref="taxon:85962"
complement(142. .519)
gene="HP1199"
complement(142. .519)
CDS /note="similar to SP:P05392 percent identity: 65.04;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L7/L12 (rpL7/L12)"
/protein_id="AAD08245.1"
/db_xref="GI:2314361"
/translation="MAISKEVLEYIGSVLESELYMFKKRGVSTPYNVAGAA
VAGGAASEKEETEFNVILADSGAEKIKIVKVRBITGLKEADATEKTPHVLKEG
VNKEAETIKKLEEVGAKVEVK"
complement(565. .1059)
gene="HP1200"
complement(565. .1059)
CDS /note="similar to SP:P42923 PID:786163 GB:AL009126 percent
identity: 30.38; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L10 (rpL10)"
/protein_id="AAD08246.1"
/db_xref="GI:2314362"
/translation="MQKQHQKHVELVANLKSOPADAKALLICDKGLSVKLEBALR
NKARNOGKVOYIKNTLAHIAKETSGLDKETNPLMGSDOIALSKLVDPDKKH
KDPFLKXGLPEPKESVSAHVAVSKLPSKEELMGLSVWTAAPARYPTGTDLNRKA
KEEN"
complement(1168. .1872)
gene="HP1201"
complement(1168. .1872)
CDS /note="similar to GB:LA2023 SP:P44342 PID:1003916
PID:1222450 PID:1204767 percent identity: 51.98;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L1 (rpL1)"
/protein_id="AAD08247.1"
/db_xref="GI:2314363"
/translation="MAKVFKELEKLFKIQNDKAYGVQGVVKSLSAKFDEIVE
VALRLGVDPRHADQWAGAVVLPHTGKKVAVAPAKDIKQDEANAGADVGGDLA
EIKKGRIDFDMVIAIPDMMAVAVGRLGKGLPMPKGTVMIDIAKAVSNKSG
GVNPRVDKGNVHARIGKASPEEKIKENMELVVTIRLAKSSAKGTYRNAALSLT
MSPSYLDAQELMDIK"
complement(1917. .2342)
gene="HP1202"
complement(1917. .2342)
CDS /note="similar to SP:P29395 GB:211839 PID:48184
GB:AE000512 percent identity: 63.83; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L11 (rpL11)"
/protein_id="AAD08248.1"
/db_xref="GI:2314364"
/translation="MAKRVGEIKLQIPAGKANPSPVGPALIGRGVIMERCFAFNE
RTKMGSPNIPVITIVDQSEFTFKKPPVLDLKKASGVKSGSDNPLKNNIATLTH
KQVEBIADLKMDLNTSTWEAKKIVMGASRMGEVVD"
complement(2360. .2890)
gene="HP1203"
complement(2360. .2890)
CDS /note="similar to SP:P16921 PID:147801 PID:396321
GB:U00096 PID:1790414 percent identity: 41.04; identified
by sequence similarity; putative"

gene /codon_start=1
/transl_table=11
/product="transcription termination factor NusG (nusG)"
/protein_id="AAD08249.1"
/db_xref="GI:2314365"
/translation="MMDWYAIQTSYSGSSQYKKAIENTLANDHNIDRIQELIIVPEDI
IEVSKSKTKYTERSLPGYVFIKVDLDTVMHKIQSLPRVSRFGENKKPTPLSAD
IGHILEKNNRANAPKPKIFPEQGEVVRVEGFPANFTATVEEDVDEHRLKLNVSIFG
RNTPEILHSQVEKII"
complement(3242. .3314)
gene="CRNA-Tip-1"
/gene="CRNA-Tip-1"
complement(3242. .3314)
CDS /note="CRNA-Tip-1"
/product="CRNA-Tip-1"
/codon_start=1
/transl_table=11
/product="ribosomal protein L33 (rpL33)"
/protein_id="AAD08255.1"
/db_xref="GI:2314371"
/translation="MKVIGLKSCDCEIDINVTGNKNTKEKLKKFCRENCHTL
HKEIKLKS"
complement(3548. .4747)
gene="HP1205"
complement(3548. .4747)
CDS /note="similar to SP:P42482 PID:587590 percent identity:
89.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="translacion elongation factor EF-Tu (tufB)"
/protein_id="AAD08250.1"
/db_xref="GI:2314366"
/translation="MAKKEFRNTRKHVNIIGTGHVDKGTTSAAISAVLSKGLAEM
KDYDINDIAPBEKERGITATSHLEYETENRYAVVDCPGHADVYKMMITGAQNDGA
ILVVSADGPMPOTEHILSRQVPHIVFLNKQDWDQDELLEVMERYBLISA
VEPFGDDPIIVAGSLRALREBAKAGNVEMGSKVLAAYDAVYIPTBERDYEKFLM
PVEVDFSIAGRTVATGRIERGIVVGVDEIVEIGIRPTOKTGVEMRKRLEKGBA
GDNVGVLRGKTKKEVERGCMVLCRGSITPHKKFGELIYVLSKEGCHHTPEPTVYRP
QYVRTIVTGSITLPEGVEMVPEQDNKITYELISPALELGTFAIRREGRTVAG
VSNIIIE"
complement(4877. .4948)
gene="CRNA-Tyr-1"
/gene="CRNA-Tyr-1"
complement(4877. .4948)
CDS /note="CRNA-Tyr-1"
/product="CRNA-Tyr-1"
/codon_start=1
/transl_table=11
/product="CRNA-Tyr-1"
/note="codon recognized: ACC"
complement(4982. .5055)
gene="CRNA-Gly-2"
complement(4982. .5055)
CDS /note="CRNA-Gly-2"
/product="CRNA-Gly-2"
/codon_start=1
/transl_table=11
/product="CRNA-Gly-2"
/note="codon recognized: GGA"
complement(5074. .5155)
gene="CRNA-Tyr-1"
complement(5074. .5155)
CDS /note="CRNA-Tyr-1"
/product="CRNA-Tyr-1"
/codon_start=1
/transl_table=11
/product="CRNA-Tyr-1"
/note="codon recognized: UAC"
complement(5186. .5258)
gene="CRNA-Tyr-2"
complement(5186. .5258)
CDS /note="CRNA-Tyr-2"
/product="CRNA-Tyr-2"
/codon_start=1
/transl_table=11
/product="CRNA-Tyr-2"
/note="codon recognized: ACA"
5448. .7184
/gene="HP1206"

CDS

5448..7184
/gene="HP1206"
/note="similar to SP:P22638 PID:142020 percent identity:
26.18; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="multidrug resistance protein (mefA)"
/protein_id="AAD08251.1"
/db_xref="GI:2314367"
/translation="MAKKKHKSTLKYLRLSKQIYMLTFEKRVVFLVLMVAFSS

Alignment Scores:

Pred. No.:	8.3e-127	Length:	10532
Score:	2043.00	Matches:	399
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-10-039-183a-4 (1-399) x AEO00626 (1-10532)

QY 1 MetAlaLysGluLysPheAsnArgThrIysProHisValAsnIleGlyThrIleGlyHis 20

DB 4747 ATGGCAAAAGAAAGTTTAAACAGAACTAAGCCGCAATTAATTAATGGAACCAATGGGCAT 4688

QY 21 ValAspHisGlyLysThrThrLeuSerAlaIleSerAlaValLeuSerLeuLysGly 40

DB 4667 GTAGACCATGTAAACGACTTTGAGTGCAGCGATTTTCACCGCTGCTTTCTTGAAGGT 4628

QY 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60

DB 4627 CTTCGCAAAATGAAAGACTATGATTAATGATTAACGCCCTGAAAGAAAGAAAGAGCG 4568

QY 61 IleThrIleAlaThrSerHisIleGluLysGluThrGluAsnArgHisTyrAlaHisVal 80

DB 4567 ATCACTATGCTACTCTCTACATGTGAATGTAGACTGAAAGACAGACTATGCGCATGTG 4508

QY 81 AspCysProGluLysAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGluMetAsp 100

DB 4507 GATTCCCGACGACACGCTGACTATGTAAACATGATACCGGTCGCGCAATTGAC 4448

QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGluThrArgLysIle 120

DB 4447 GGAGCGATTTGGTTGTTCTGACGTGAGCCCTCATCCCTCAAACTGAGGACATATC 4388

QY 121 LeuLeuSerArgGluValGlyValProHisIleValIlePheLeuAsnLysGluAspMet 140

DB 4387 TTATGTCTCGTCAAGTAGCGTGCCTCATCGTGTGTTCTTAAACAAACAGACATG 4328

QY 141 ValAspAspGluGluLeuLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160

DB 4327 GTAGATGACCAAGATTTGTTAAGCTGTGAGAAATGGAAGTCCGCAATTTTGAGCGCG 4268

QY 161 TyrGluPheProGlyAspAspThrProIleValIleGlySerAlaLeuArgAlaLeuGlu 180

DB 4267 TATGATTTCTCGCGCATGACACTCTATCGTAGCGGTTTCACTTAAGACCTTAGAA 4208

QY 181 GluAlaLysAlaGlyAsnValGlyGluTyrPglYglLysValLeuLysPheLeuMetLaglu 200

DB 4207 GAAGCAAAAGCTGTGAATGTGGTGAATGGGTGAAAGAGTCTTAAACTTAATGCTGAA 4148

QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220

DB 4147 GTGGATGCTATATCCCTATCTCAGAAAGACATGAAAGAACTTTCTTGATGCCGGTT 4088

QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240

DB 4087 GAAAGATGTCTCTATGCGGGTAGAGGAGCTGTGTATACAGATGAGATTAAGAGAGCG 4028

QY 241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGluLysThrThr 260

DB 4027 GTGTGAAGATGAGCGATGAAAGTGAAGATCGTTGTATCAGACTTAACAAAGAAAGACT 3968

QY 261 ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal 280

DB 3967 GTAAACCGGTGTAAGAATGTTTAGAAAGAGTTGAAAGAGTGAAGCCGCGCATATATGTG 3908

QY 281 GlyValIleLeuLeuArgGlyThrIleLysGluGluValGluLysGlyMetValLeuLysLys 300

DB 3907 GCGGTGCTTTTGAGAGGAACCTAAAGAAAGAGTGAACCGCGTATGTGTTATGCAAA 3848

QY 301 ProGlySerIleThrProHisLysLysPheGluGluGluIleTyrValLeuSerLysGlu 320

DB 3847 CCAAGTTTATACACCTCCGACAAAGAAATTTAGAGGAGAAATTTATGCTCTTAAAGA 3788

QY 321 GluGlyGlyArgHisIleThrProPheThrAsnTyrArgProGluPheTyrValArgThr 340

DB 3787 GAAGCGGAGAGACACATCTCAATTTCTTCAACCAATTCACCCGCAATTCATGCGCACA 3728

QY 341 ThrAspValThrGlySerIleThrLeuProGluGluValGluMetValMetProGlyAsp 360

DB 3727 ACTGATGTGACTGGCTTATACCCCTTCCTGAAGGCGTGAAGATGTTATGCTGCGCAT 3668

QY 361 AsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrLysPhe 380

DB 3667 AATGTGAATATCACTGTAGATGATTAAGCCCTGTGCGCTTAGACTTGGAACTAAATTT 3608

QY 381 AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu 399

DB 3607 GCGATTCGTGAAGGCGGTAGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGAA 3551

RESULT 3

BD009865

LOCUS

DEFINITION

Proteins, in particular membrane proteins, of Helicobacter pylori, their preparation and use.

ACCESSION

BD009865

VERSION

BD009865.1 GI:18638238

KEYWORDS

JP 2001502886-A/6.

SOURCE

ORGANISM

Helicobacter pylori

REFERENCE

1 (bases 1 to 2825)

AUTHORS

Knapf, B., Hundt, B. and Schmidt, K.H.

TITLE

Proteins, in particular membrane proteins, of Helicobacter pylori, their preparation and use

JOURNAL

Patent: JP 2001502886-A 6 06-MAR-2001;

CHIRON BEHRING GMBH & CO

COMMENT

OS Helicobacter pylori

PN JP 2001502886-A/6

PD 06-MAR-2001

PF 25-JUL-1997 JP 1998508651

PI 25-JUL-1996 DE 196 30 390.7

PI BERNHARD KNAPF, ERIKA HUNDT, KARL HEINZ SCHMIDT PC

C12N15/31, C07K14/205, C07K16/12, G01N33/53, A61K31/70, A61K39/106, PC

A61K39/395

CC Strandedness: Single;

CC Topology: Linear;

FT source

FT 1..2825

location/Qualifiers

1..2825

/organism="Helicobacter pylori"

/mol_type="genomic DNA"

/db_xref="taxon:210"

ORIGIN

Alignment Scores:

Pred. No.:	7.46e-127	Length:	2825
Score:	2035.00 <td>Matches:</td> <td>398 </td>	Matches:	398
Percent Similarity:	99.75% <td>Conservative:</td> <td>0 </td>	Conservative:	0
Best Local Similarity:	99.75% <td>Mismatches:</td> <td>1 </td>	Mismatches:	1
Query Match:	99.61% <td>Indels:</td> <td>0 </td>	Indels:	0
DB:	6 <td>Gaps:</td> <td>0 </td>	Gaps:	0

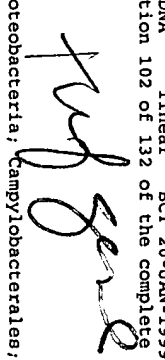
US-10-039-183A-4 (1-399) x BD009865 (1-2825)

Oy 1 MetalAlvBgluVpheaBmaRgThrlySP-roHleValaenilegLyThrlleGlyHs 20
 Db 891 ATGGCAAAAGAAAGATTAAACGAACTAAGCCGCACTTTAAATATGGAACCAATGGGCAAT 950
 Oy 21 ValaBPHleGlyVpThrleuSerAlaAlaileSerAlaValleuSerleuVgLy 40
 Db 991 GTAGCCATGTGTAACCACTTGAAGTGCAGCAATTTGAGCGGTCTTCTTGAAGGT 1010
 Oy 41 LeuAlaGluMetLeuVpThrleuSerAlaAlaileSerAlaValleuSerleuVgLy 60
 Db 1011 CTGCAAGAAATGAAGACTAT 1070
 Oy 61 IlleThrlleAlaThSerHleleGlyVpThrlleValaBmaRgThrlySP-roHleVala 80
 Db 1071 ATCACTATCGCTACCTCCCACTTGAATATGAGCTGAACCAAGACACTATCGCATGTG 1130
 Oy 81 AepCySProGlyHleAlaBpThrlyVpThrlleValaBmaRgThrlySP-roHleVala 100
 Db 1131 GATTCGCCAGACAGCTGACTAT 1190
 Oy 101 GlAlaAlaileuValaValSerAlaAlaBpGlyPromeProGlnThrArgGluHle 120
 Db 1191 GGAGCGATTTGGTGTCTTCTGACGTATGCGCTTATGCTCAACCAAGACACTATAT 1250
 Oy 121 LeuLeuSerArgGluValaGlyVpThrlleValaBmaRgThrlySP-roHleVala 140
 Db 1251 TTATGTCTCGCTCAAGTGAAGTGCCTGCTCACTGTTGTTCTTAAACCAACCAAGT 1310
 Oy 141 ValaBpArgGluLeuLeuGluLeuValaGluMetGluValaArgGluLeuLeuSerAla 160
 Db 1311 GTAGTGAACCAAGATGTAT 1370
 Oy 161 TyrGluPheProGlyVpThrlleValaBmaRgThrlySP-roHleVala 180
 Db 1371 TATGATTCCTCGTGAAGTGCCTGCTCACTGTTGTTCTTAAACCAAGT 1430
 Oy 181 GlAlaAlaileuValaValSerAlaAlaBpGlyPromeProGlnThrArgGluHle 200
 Db 1431 GAAGCAAAAGCTGTGTAATGCTGCTCACTGTTGTTCTTAAACCAAGT 1490
 Oy 201 ValaBpArgGluLeuLeuGluLeuValaGluMetGluValaArgGluLeuLeuSerAla 220
 Db 1491 GTGATGCTCAATATCTTCACTCAAGAAAGACACTGAAACCTTCTGATGCGGT 1550
 Oy 221 GluBpValPheSerAlaAlaileSerAlaValleuSerleuVgLy 240
 Db 1551 GAAGTGTCTTCTTATGCGGTGAGGCACTGTGTTACAGTGAAGTGAAGT 1610
 Oy 241 ValValleuValaGlyVpThrlleValaBmaRgThrlySP-roHleVala 260
 Db 1611 GTGCGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1670
 Oy 261 ValThrglyValaGluMetPheBpGlyVpThrlleValaBmaRgThrlySP-roHleVala 280
 Db 1671 GTAGCGGTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1730
 Oy 281 GluValleuLeuValaGlyVpThrlleValaBmaRgThrlySP-roHleVala 300
 Db 1731 GCGGTGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1790
 Oy 301 ProGlySerleThrProHleValaBmaRgThrlySP-roHleVala 320
 Db 1791 CCAGTGTCTTCACTCCGCAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1850
 Oy 321 GluGlyVpThrglyValaGluMetPheBpGlyVpThrlleValaBmaRgThrlySP-roHleVala 340
 Db 1851 GAAGCGGAGACACACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1910
 Oy 341 ThrAepValThrglySerleThrleuPProGlyValaGluMetValaBpProGlyVp 360

Db 1911 ACTGATGACTGCTTATACCTTCTGAGGCGTGAAGGCTTAATGCTGCGCAT 1970
 Oy 361 AsnValleuSerleThrValaGluLeuLeuSerProValaAlaileuGluLeuGlyThrllePhe 380
 Db 1971 AATGTGAATCACTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2030
 Oy 381 AlaileuSerleuGlyVpThrglyValaGluMetPheBpGlyVpThrlleVala 399
 Db 2031 CGGATTCGTGAAGGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2087

RESULT 4
 AE001541
 LOCUS
 DEFINITION Helicobacter pylori, strain J99 section 102 of 132 of the complete genome.
 ACCESSION AE001541
 VERSION AE001541.1 GI:4155724
 KEYWORDS
 SOURCE
 ORGANISM Helicobacter pylori J99
 Helicobacter pylori J99
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.
 REFERENCE
 AUTHORS
 1 (bases 1 to 15292)
 Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Carnese, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F. and Trust, T.J.
 Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
 Nature 397 (6715), 176-180 (1999)
 JOURNAL
 MEDLINE
 PUBMED
 9923682
 2 (bases 1 to 15292)
 King, B.L., Alm, R.A. and Trust, T.J.
 Direct Submission
 Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA
 COMMENT
 Address all correspondence to: hph@arch.us.astra.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Brydon C. Guild, Gilles Carmel, Anthony Carnese, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: http://www.astra-boston.com/hpylori).
 Location/Qualifiers
 1..15292
 /organism="Helicobacter pylori J99"
 /mol_type="Genomic DNA"
 /strain="J99"
 /db_xref="taxon:85963"
 complement(257..415)
 /gene="rpmg"
 /note="synonym: jhp1127"
 complement(257..415)
 /gene="rpmg"
 /note="similar to H. pylori 26695 gene HPI204"
 /codon_start=1
 /transl_table=1
 /product="50S RIBOSOMAL PROTEIN L23"
 /protein_id="AAD06710.1"
 /db_xref="gi:4155730"
 /translation="MKVKIGLKSCDIEDINVTTKAKTTEKLEKFCPRENHGTL HKKIKLKS"
 complement(462..1661)
 /gene="tufA"

FEATURES
 source
 gene
 CDS
 gene



gene
CDS
/note="synonym: jhp1128"
/complement(462..1661)
/gene="tufa"
/note="similar to H. pylori 26695 gene HP1205"
/codon_start=1
/transl_table=11
/product="ELONGATION FACTOR TU (EF-TU)"
/protein_id="AAD06711.1"
/db_xref="GI:4155731"
/translation="MAKEFNPTNPHNIGTIGHYHCKTTLSAISAVALSISKLAEM
KDYNDINAPQEKGITITATSHISETETRTYAHVDCGHADYKNNITGAQNDGA
ILVVSADGPMQETREHILSRQGVPHI VFLNQDWDQDELLELVEMERELLISA
YEPFGDDPIVAGSALRALBESAKQNVGSGKVLMAEVDSYLPTERDYKTFILM
FVEDVFSIAGRTVTVGRIERGVVAGDEVLVGRATQKTGTGEMFRKLEGEA
GDNVGLLRGTKEVERGMVLCRPGSITPHKPFEEIYVLKSEGGRRHTPFTYRP
QFVYPTVTGSIITLPEGVEMVPMQDNKITVELISVALBELGTKFAIREGRTYAG
VVSNIIE"
2365..4101
/gene="jhp1129"
2365..4101
/gene="jhp1129"
/note="similar to H. pylori 26695 gene HP1206"
/codon_start=1
/transl_table=11
/product="abc transporter, ATP-binding protein"
/protein_id="AAD06712.1"
/db_xref="GI:4155732"
/translation="MAKKKKIKPTLKPLRSIKOIVRLITPEKMIFFPLVIMAVSS
FVEMSLTILMPFITLADPSALDDKXMKYDFFHSPSPRLMFPSPFCVGYILF
RMFGVFFTYLGRFSNKKAYOIKOOLPLOHKNNYLNNHNLDSLDIINNKKEGM
EMSFNAPLSLLEIVIVFYSTLILTNWKTILVFTTILALQIFPIVKRVVLIRKKG
EMAAKSAQOTLKEIVSFNSFKITLKDNEBAHLFGNSRKADTEIIVSTLQVVP
RYSIETVGSLLILAAVILFPGYGEARVLPISVALALYVILPSVTGVSYYEIA
YNOLATNVFSLISKTIVBEDVLDPNFKITLQNIISPAVYSKHPVLNPNLTIOKG
KIALIGHSGCCSTIADITMGLTTPKSGEIPFDNTLITSENRSRKIKGTIPQITYL
FDGTVDGNIARFSAIDERLIVCRMHHYDLCEHGLKTQVGEAKLSGQKQRI
GIARALYDNPETLVIDEATSLDNETESKINDEIYQIAKNKTLIVIAHRLSTIERCEV
IIDMSQKDNLS"
complement(4110..4778)
/gene="jhp1130"
complement(4110..4778)
/gene="jhp1130"
/note="similar to H. pylori 26695 gene HP1207"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06713.1"
/db_xref="GI:4155733"
/translation="MALEVVLMDPDGVI FDSNMLKYEGFKALFQKHGNSKEGLKQPE
VYHYSOGISRNKEIKQYFNEILKTPIAQEBIDALAFGAILBOKLFDRGHLNSEVM
AFIDKHQNYIFHISALHSELVLCSEFLGTTKYFKSVESGPPDKPIIANIIQKVA
YDPSRMILIGDSVNDYESAKANKVAFLGNSKVLKNLVGDQYQGYLESFPGPLQN
PAKE"
complement(4966..5955)
/gene="M.Hpy1"
/note="synonym: jhp1131"
complement(4966..5955)
/gene="M.Hpy1"
/note="similar to H. pylori 26695 gene HP1208"
/codon_start=1
/transl_table=11
/product="TYPE II DNA MODIFICATION ENZYME
(METHYLTRANSFERASE)"
/protein_id="AAD06714.1"
/db_xref="GI:4155734"
/translation="WNYIGSKYKLIPIKENIHAVAHDLSAIFCDLPACTGIVGRT
FKKAVNKVINDLEYEYFVLNQNYIGNIOEIPNKEELINKINSVALKKGFIYSHSLG
GSSROYFSETNAOKIDAMRLKLEBLKSONIDNHXYFLASLGSADKAVNATAVYG
AFIKRLKSAOKELIKGAHPVSLNANEBVDSNDIGKISGILLVLDPPYARQY
GANHLIANTIAATPTPKGKTDLPSYOKSSGCSFQILNAPENIKKARPRYITLSTY
NNEGIMSETIKKYGAYSIVTKTYMRFPADNKRRAHKAHVHKECLHVLIK"
complement(6284..6463)
/gene="jhp1132"

gene
CDS
complement(6284..6463)
/gene="jhp1132"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06705.1"
/db_xref="GI:4155725"
/translation="MAAVIKVNGKIQEYENGNYKRTYDSNIVAADTGHIVAAYTAK
GKVEEYKNGIKHRTY"
complement(6692..7207)
/gene="cysB"
/note="synonym: jhp1133"
complement(6692..7207)
/gene="cysB"
/note="similar to H. pylori 26695 gene HP1210"
/codon_start=1
/transl_table=11
/product="O-SERINE ACETYLTTRANSFERASE"
/protein_id="AAD06715.1"
/db_xref="GI:4155735"
/translation="MLDSYSLERVLQDPAPARKMEVLLYPGIIHALCYRLAHALH
KRPFYIARALSOLARPIYTGIEIHGAKIGGLFIDHGAGVIGETTEIGDVTIYHG
VTLGTGKFKRKRHPTLGNRVVAGATVLAICVGDVIRIGANNAVLSDLPTGSTAV
GAKATITXDR"
7507..7602
/gene="jhp1134"
7507..7602
/gene="jhp1134"
/note="similar to H. pylori 26695 gene HP1211"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD06716.1"
/db_xref="GI:4155736"
/translation="WVPKVGLEPRDSCSHQILSLACLPPLRURY"
complement(7804..8121)
/gene="atpB"
/note="synonym: jhp1135"
complement(7804..8121)
/gene="atpB"
/note="similar to H. pylori 26695 gene HP1212"
/codon_start=1
/transl_table=11
/product="ATP synthase F0, subunit c"
/protein_id="AAD06717.1"
/db_xref="GI:4155737"
/translation="MKPLATPFLIAGVAFPAHDGSGMDIKSYSLIGAMIGIGIAA
FGATGKGNAAAATITGTARNPGVGKLLITTFVAMANTEAVQVITVFAIATYSNP
FLS"
complement(8254..10320)
/gene="pnp"
/note="synonym: jhp1136"
complement(8254..10320)
/gene="pnp"
/note="similar to H. pylori 26695 gene HP1213"
/codon_start=1
/transl_table=11
/product="POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE"
/protein_id="AAD06718.1"
/db_xref="GI:4155738"
/translation="MDPITINSGNRTTEPALQYAKQATSSLMRLGKTLILASVCBE
REPVSEDLPLVVOFLKSYAAGAKIPGSPVVRGEAODFEILSLIDRTLRLPLPKD
RYRPIQITLMTLVNHDLENDLVASALNAAATLFLAHLPIKYSACRLARVDNEFIIN
PSALINOSSLDLPVSGKESLMTMEMSISGOKNALAEPLMLALAELOAKSKEKCA
LYEARFTPYQNELLPKSEBGIVLNERLLDLKNQYFDITIKIGIBSALSBRNNYKEV
AKTISEAHSEPSLEIBLSTLSEKVKTEIRRM10DKIRPDYRALBEVP1STESNLLP

Alignment Scores:
Pred. No.: 7,786-124 Length: 15292
Score: 2001.00 Matches: 392
Percent Similarity: 98.75% Conservative: 2
Best Local Similarity: 98.25% Mismatches: 5
Query Match: 97.94% Indels: 0

DB: 1 Gaps: 0

US-10-039-183a-4 (1-399) x AE001541 (1-15292)

QY 1 MetAlaIyGulUyPheAsnArgThrybProHlvaIAsnIleGlyThrllegIyHis 20

DB 1661 ATGGCAAAAGAAATTTATAGAACTAACCCGCACTGTTAATATTGAAACCAATGGGCAT 1602

QY 21 ValAspHlAglyLyThrlThrlSerHlAaIleSerAlaValLeuSerleuylsgly 40

DB 1601 GTATACATGTGTAACGACTTGGAGCGGCAATTTCCGCGTCTTCTTGAAGGT 1542

QY 41 LeuAlaGluMetLeuAspTyAspAsnIleAspAsnAlaProGluGluylsgly 60

DB 1541 CTTCGAGAAATGAAAGACTATGACAAATNTGATTAACGCCCTCAGAAATAAGAGAGG 1482

QY 61 IlleHrlleAlaThrlSerHlIleGlyThryGluThryGluAsnArgHlAaIle 80

DB 1481 ATCACTATCGCTACTCTACATTTGAAATGAGACTGAAACGACACTATCGCATGTG 1422

QY 81 AspCybProGluYHlAaIAspTyValIyAspMetIleThryGluAlaIleGluMetAsp 100

DB 1421 GATTGCCAGAGACAGCTGACTATGTAAACATGATCACCAGTGGCGCAAAATGAT 1362

QY 101 GluAlaIleLeuValIleSerAlaIleAspGlyPrometProGluThryArgGluHisIle 120

DB 1361 GGAGGCACTTGTTGTTCTTCCGCGTATGAGCCCATGCTCAACCAAGAGAGCATATC 1302

QY 121 LeuLeuSerArgGluValIyGlyProHlIleValIlePheLeuAsnlysglnAspMet 140

DB 1301 TTATGTCTCGTCAAGTAGGCGGTGCTCATGTTGTTTCTTAAACAAACAGACATG 1242

QY 141 ValAspArgGluGluLeuGluGluValIleGluMetGluValIyGluLeuSerAla 160

DB 1241 GTAGATGACCAAGAAATGTTAGAGCTTGTAGAAATGGAAGTACGAGTGTGGCGCG 1182

QY 161 TyrGluPheProGluYAspAspThrlProIleValIleGlySerAlaLeuArgAlaLeuGlu 180

DB 1181 TATGAAATTCCTGGCGGATGACCTCTATCGTAGCGGGTTCAGCTTTAAGACTTTAGAA 1122

QY 181 GluAlaIyAlaIyAsnValIyGluThryGlyGluYValLeuLyLeuMetAlaGlu 200

DB 1121 GAAGCAAAAGCTGCAATGCTGATGATGCGGCGAAAGGCTTAACATCATGCTGTA 1062

QY 201 ValAspAlaIyThrlleProThrlProGluArgAspThryGluYThrlPheLeuMetProVal 220

DB 1061 GTGGATTCCTATATCTTACTCCAGAAAGACACTGAAATACTTCTTGTATGCGGAT 1002

QY 221 GluAspValIlePheSerIleAlaIyArgGlyThryValIleThryGlyArgIleGluArgGly 240

DB 1001 GAAGACGTGTTTCTATCGCAGGTAGAGGAACTGTGTTCAGGTAGATTGAAAGAGGC 942

QY 241 ValValIyIyValIyAspGluValIleGluIleValIyIleArgProThrglnlyThrlThr 260

DB 941 GTGGGAAAGTGGCGAGTGAAGTGAATCGTTGGATCAGAGCCACCAAAAAAGCACC 882

QY 261 ValThryGlyValIleGluMetPheArgGlyLeuGluGluYValIyGluAlaIyAspAsnVal 280

DB 881 GTAAACCGGTGTAAGAAATGTTAGAAAGAAATTAAGAAAGGTGAAGCCGCGCATATGTG 822

QY 281 GluValIleLeuLeuArgGlyThryIyIyIyGluGluValIyGluArgGlyMetValIleuCylys 300

DB 821 GCGGTCCTTTGAGAGGAACTAAAGAAAGAAAGTGAAGCCGCGGTATGCTTATAGCAAA 762

QY 301 ProGlySerIleThrlProHlIyIyIyPheGluGluIyGluIleThryValLeuSerIyGlu 320

DB 761 CCAAGTTCTATCACTCCGCAAGAAATTTGAGAAATTTATGCTCTTCTTAAAGAA 702

QY 321 GluGlyGlyArgHlIleThrlProPheThrlAsnTyThryGluPheThryValArgThry 340

DB 701 GAAGGCGGAGACACACTTCATCTTACCAACTACCGCCCAATTCATATGCGTACG 642

QY 341 ThrAspValThryGlySerIleThrlleuProGluGluValIleGluMetValMetProGlyAsp 360

DB 641 ACTGATGACTGCTATACACCTTCCTGAGGCGTGAAGATGATGATGCTGGCAT 582

QY 361 AsnValIyIleThryValIleGluIleSerProValAlaLeuGluLeuGlyThryIyPhe 380

DB 581 AATGAAATCAACCGTGTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 522

QY 381 AlaIleArgGluGlyArgThryValIyAlaIyValIleSerAsnIleIleGlu 399

DB 521 GCGATTCGTGAAGCGGAGAAACGTTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 465

RESULT 5
AE017145/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS
TITLE
JOURNAL

521 GCGATTCGTGAAGCGGAGAAACGTTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 465

AE017145 308929 bp DNA linear BCT 26-JUN-2003
Helicobacter hepaticus ATCC 51449 section 2 of 6 of the complete genome.
AE017145 AE017125
AE017145 GI:32261855
Helicobacter hepaticus ATCC 51449
Helicobacter hepaticus ATCC 51449
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.
1 (bases 1 to 308929)
Suerbaum, S., Josenhans, C., Sterzenbach, T., Drescher, B., Brandt, P., Bell, M., Droge, M., Fartmann, B., Fischer, H. P., Ge, Z., Horster, A., Holland, R., Klein, K., Konis, J., Macko, L., Mend, G. L., Nyakatura, G., Schauer, D. B., Shen, Z., Weber, J., Frosch, M., and Fox, J. G.
The complete genome sequence of the carcinogenic bacterium Helicobacter hepaticus
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7901-7906 (2003)
22709201
12810954
2 (bases 1 to 308929)
Drescher, B. and Suerbaum, S.
Direct Submission
Submitted (27-May-2003) Institute of Hygiene and Microbiology,
University of Wuerzburg, Josef Schneider Str. 2, Wuerzburg D-97080,
Germany

FEATURES
source
Location/Qualifiers
1..308929
/organism="Helicobacter hepaticus ATCC 51449"
/mol_type="genomic DNA"
/strain="ATCC 51449"
/db_xref="ATCC:51449"
/db_xref="taxon:235279"
complement(76..996)
/locus_tag="HH0308"
complement(76..996)
/locus_tag="HH0308"
/note="GTPase, ERA family"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
protein_id="AAP76905.1"
/db_xref="GI:32261855"
/translation="MQIESISNQTPTAGPVATIGRRPNACKSTLNLNLEGNALVS
HKAVNRKQQLIVFAPLNAQILFVDPGHHQKLLNQTLSQALASLDYDALY
LAPVSDIDGHEHFLSOGKHLILSKIDCTNEILHKIAQYRSDTFLPLPV
SVKGFDEKELHTLASHLPSPFLYDEILSNLSKEIYKEMIQSLFENISDEIPY
EADVVIKKEESAEODRIFAHIIYKESQKGVVIGKGSSTIKRISKARQKIESLSK
AIFLKLDVYKMGSKORDTLKKIGYDTL"
complement(1014..2348)
/gene="hslU"
/locus_tag="HH0309"
complement(1014..2348)
/gene="hslU"
/locus_tag="HH0309"
/codon_start=1
/transl_table=1
/product="ATP-dependent protease subunit hslU/clpY"
protein_id="AAP76906.1"
/db_xref="GI:32261857"

gene
/translation="MKNMTPKQIVEYLDGYIIIGNDAKAVALAENRRYRRLKDKK
VODEITPKNIMLIGSTGVKTEIARRMAKMGLPVYKVASKYTEVGVGVDSVETK
DLVLSVNLVENEHROAQAQINDYIETIKQKLPPLSPGVSEAKSEYDMSFTKMK
ORVIGEVDDMIETIEISKKPISSPDENAMPDMIKVOESIFKVULNODKIKKMSV
KRAKALVAEIESVLDASIRAEGRRAQENGVIPIEIDVANGSDGSRQSDSKS
GVORLPLVEGSSVNTYRGQIKTDYIILPIAAGAHLSPDLPELQGFPLRVELS
MLDESLYQILQTSKSLRQYQLDLDEDIRLESDERLARLSHNAQRTDIDIG
ARLHTEIERVLEDISFDVDTYKKEVEITGDMVRERLGDVENIDLARVYL"
complement (2362. .2904)
/gene="hslv"
/locus_tag="HH0310"
complement (2362. .2904)
/gene="hslv"
/locus_tag="HH0310"
/codon_start=1
/transl_table=11
/product="ATP-dependent protease subunit HslV/CtpQ"
/protein_id="AAP76907.1"
/db_xref="GI:32261858"
/translation="MFEATTLIGYKGEYNGKCAKIIIGDQGYTFQNCVLYKNATKIRT
LYNGEILSGFAGSTADAFSLFDMFERILEGRGDLVRSVLEFSKEMRKOKYRLKEAM
MIVLNTHEIYILSGDVVEPEDGTIAAIGSGNVALAARALHNYASLPPREIVEHS
LAIAEGLCTYTNKIKIEL"
complement (2919. .3362)
/gene="tp11"
/locus_tag="HH0311"
complement (2919. .3362)
/gene="tp11"
/locus_tag="HH0311"
/codon_start=1
/transl_table=11
/product="ribosomal protein L9"
/protein_id="AAP76908.1"
/db_xref="GI:32261859"
/translation="MKVILLVNGSLGKKGEIVYKDGCGNPLAKGKAQATBEVI
NKYKQQRKABSLAEKEMQMKRTIESLTLAKKVGANDTLFGSIYEEIALAL
BEHKVSLDKHIEIAQAIKHGNGFEVQVLGGINATLKLEIKAQ"
complement (4112. .4112)
/locus_tag="HH0312"
3384. .4112
/locus_tag="HH0312"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAP76909.1"
/db_xref="GI:32261860"
/translation="MRCETITSGNIMKALKDYSYPARAHIGHKAIKYPRLIKAD
LSLEFVPLRFPWOLQAFIRKHKMIESLTKLDCNAHTQSILOASDELITGSM
QRLSLTSPFSHAOSLTLSPQNLKAPKSLIDYICSRAPDIAATMGVOYQSIKITN
ALTRGSCITYNRLFFSFMILFAPKELIDYVLIHBLAHIRYKMSQDWFDFQRCYCN
AKTRHQAQRKAKFPYALLORLSL"
complement (4120. .5073)
/locus_tag="HH0313"
complement (4120. .5073)
/locus_tag="HH0313"
/note="predicted secreted protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP76910.1"
/db_xref="GI:32261861"
/translation="MKGLIKVIYVGFAPFELVANDVDSTYFKLLSGCMMNGQLNGKT
OLCGNNDAAGSVAAYKILKPLSLSGEIVANIEPSECSDNETWSIFYPKRSQFSF
LOTARLSVADKDIOIENLIPKWPFLDIPKRPVSPSREINRERKRVIPDDDSAGNIV
PHSDIVAKYERHVSQSGSYDKINVSNTITKIKSFKPSNRDAZYENMLGUTGYSIV
TDFVNLKEITNGKLERIYTKDKEDVLIIGYNTKRSWQYFIDKKSFNANDKWK
VLVEPPRIKDVHAKAIGYIHTSQIKMFEDGAEKRLRLRGK"
complement (5083. .6687)
/gene="trpD"
/locus_tag="HH0314"
complement (5083. .6687)
/gene="trpD"
/locus_tag="HH0314"

gene
/EC_number="2.4.2.18"
/codon_start=1
/transl_table=11
/product="anthranilate phosphoribosyltransferase"
/protein_id="AAP76911.1"
/db_xref="GI:32261862"
/translation="MILIDYDFTYNIYQAFYRFGPPIRVARSDKISIEIRELNP
OYIILIGPQPKPKAGISIQIVQFQGYPIIGICLGHQAUMAAFQMDIIVAKNTANG
KVEPLHNOKLPHRITPLTPYVRSVHSGANLEPDCFEVSAMEDEIWAIEKHII
HLVQVPHPEISIGTGEKMLNPLHATREIIPITKQYKSTLKGENSLFKQAYVMD
ITEGNMSDAQIGSITLSLEIKYNAEBELAGASVUKKAIAPPLPLSEVRLDWTG
GSPHKTENVSTTSALLLAAGVKIIKGNRAITSSGADLLSALGVNIHNEVQCID
VYKNGITFLFPAQKFAHMPAPARATLGRTAKNLGLPLANPANTVHOFIYVDFKA
YTEIMARALDILGIRAMVVSQFDYDEISLCAPTOILELNNGVYKSYVFNPSIEIGD
FVYTELGGGVQENKRISLIDIPASPOSLSPTQVVALNTGALYIADRAQSIKEG
YFLARDIESKRVALLIEFVRLSNQ"
complement (6691. .8187)
/gene="trpE 2"
/locus_tag="HH0315"
complement (6691. .8187)
/gene="trpE 2"
/locus_tag="HH0315"
/EC_number="4.1.3.27"
/codon_start=1
/transl_table=11
/product="anthranilate synthase component I"
/protein_id="AAP76912.1"
/db_xref="GI:32261863"
/translation="MFPDSNGRISIAKKEILCNTPPLAVISINAKVLESANVT
GKDRFSYILNEAFRIYKCGAFLTYTGQKIPLYEALADYTPLDLSLPTPHKAGFL
ESLALVRLTAPPTDSIPALPLPLGGAGYGYEFPFAIEVTRKNSLVNAPCAFL
EGRDFLIFDLNPLRHIVSVGAHERBESHYKORIEHIAHSLATLPQNGDSQKSD
PECVATSSQCESESNVEAIKKAIVAGNLLQCVSQYQIIRSLSLPLQAYRMLRQNP
PMFYVYNNDDFVITGASPEIMLRKSKQGGFPLTRPIAGTRPGESVADLBETEL
LNDKENAHEMLMLDIARNDAKVSQGVGVYIANKTEYRSRPMYVSVQGLDESK
HFSRADARSVTPAGTISGARPIKAIKITESLEHSHARCVYGAIGYFYEDMDPALA
IRTVYQGMVYMSGAGVQDSIPSAEWMETQKVRSQLDMFINDE"
complement (8218. .8757)
/gene="tolA"
/locus_tag="HH0316"
complement (8218. .8757)
/gene="tolA"
/locus_tag="HH0316"
/codon_start=1
/transl_table=11
/product="outer membrane lipoproteins carrier protein
tolA"
/protein_id="AAP76913.1"
/db_xref="GI:32261864"
/translation="MEVLRVYLFTSLCMTLPAWGENILOSEADFEQIAVNEQVPT
RYEGHIIQKMSKVMYKADPLQKEIYNNKQEVIMYBRLQVSHSRDKADAFSLIL
KSAKHNDGTHTVDGIEYVLFIDKQKPRIEVDSMGVKTTLKLNKVLNGAISD
KKAFVFPQGYEIVELHSR"
9016. .11430
/gene="seca"
/locus_tag="HH0317"
9016. .11430
/gene="seca"
/locus_tag="HH0317"
/codon_start=1
/transl_table=11
/product="preprotein translocase subunit SecA"
/protein_id="AAP76914.1"
/db_xref="GI:32261865"
/translation="MQRLKLIQNETSLSQSLHKSFAITRARSRYVGMRFVQVLIG
GMAINDGRIAMKTKGEGTLVATLAVCLNACGSGVHVTYNDVLIANDAKLEPLIN
FLGFEVGIITSEVDNDRLOAYACDVIYGTNNBFGDYLLDNMKYDLSQYQGEHYF

Alignment Scores:
Pred. No.: 9 3e-116 Length: 308929
Score: 1900.00 Matches: 366
Percent Similarity: 96.24% Conservative: 18
Best Local Similarity: 91.73% Mismatches: 15

Query Match:	93.00%	Indels:	0
DB:	1	Gaps:	0
US-10-039-183A-4 (1-399) x AB017145 (1-308929)			
Qy	1	MetAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis	20
Db	57020	ATGCCAAGAAGAAAGTTTGTGAACAACAACCTCACGTAATATGTGGCAACAAATCGGCAT	56961
Qy	21	ValAspHisGlyLysThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLysGly	40
Db	56960	GTTGATCATGTAAACAACCTTGAGTCAGCTATTCTCTGCTGTGTGGCAACTAAAGGT	56901
Qy	41	LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly	60
Db	56900	CTTCAGAAATTGAAGACTATGATAACATTGATACCGCACAGAGAAAGAAAGAGGT	56841
Qy	61	IleThrIleAlaThrSerHisIleGluThrGluAsnArgHisTyrAlaHisVal	80
Db	56840	ATTACTATCGCGACTTCTCATATTGATATGAGACAGAAATCGCCACTACGCGCACGTG	56781
Qy	81	AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGlnMetAsp	100
Db	56780	GATTGCCCGCACAGCTGACTATGTTAAATAATGATTACAGGTGCAGCTCAATGGAC	56721
Qy	101	GlyAlaIleLeuValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle	120
Db	56720	CGAGCAATTCTGTGTTCTCGCGCAGATGCTCTATGCCACAAACACGTGAGCATATC	56661
Qy	121	LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMet	140
Db	56660	TGTGTGTCGCCAAGTAGGTGCGATTACATCGTTGTTATTTTAAATPAAACAAGATATG	56601
Qy	141	ValAspAspGlnGluLeuLeuGluValGluMetGluValArgGluLeuSerAla	160
Db	56600	GTTGATGATGCTGAATTACTTGAGCTTGGAATGGAATGCGTGGAATTGCTAGTCAA	56541
Qy	161	TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu	180
Db	56540	TATGATTTCCTGGAGATGATACCTCTATCATTCAGGTCTCTGCTCTCAAGCGCTTCAA	56481
Qy	181	GluAlaLysAlaGlyAsnValGlyGluThrGlyGluLysValLeuLysLeuMetAlaGlu	200
Db	56480	GAACCAAGCAGGTAATGTAGGTGAATGGGGTGAAGAAAGTATTAAATTTAATGGAAGA	56421
Qy	201	ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal	220
Db	56420	GTTGATAAATATATTCCAACTCCACAGCTGATCTGAAAAAACAATTCCTTATGCTGTA	56361
Qy	221	GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly	240
Db	56360	GAAGATGTATTCTCAATCGCTGGTCGCAACGGTTGTAACAGGAAGAGTAGAAGAGGT	56301
Qy	241	ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThr	260
Db	56300	GTTGTGCAAGTAGCGGATGAAGTAGAAATCGTAGGAATCCGCGACACTCAAAAACAACA	56241
Qy	261	ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal	280
Db	56240	GTTACAGCGCTTGAAATGTTTAGAAGAGCTTGTAAGAGGTGAGCGCAGAGATATGTC	56181
Qy	281	GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys	300
Db	56180	GGTATTTGTTGCGGGAACAAAAAGAAAGTAGAAGAGGTATGTTCTTTGTGAAG	56121
Qy	301	ProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGlu	320
Db	56120	CCCGCTCAATTACTCCACATAAGAAATTTGAAGGTGAATTTATGTCCTTTCAAAAGAT	56061
Qy	321	GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArgThr	340
Db	56060	GAAGGTGGAACATACACCACTTTTCAATGGCTATCGCCCAACAATTTATGTGCGAACT	56001
Qy	341	ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp	360
Db	56000	ACCGATGTTACCGGTTCTTATTGAGCTTCCAAAGTGGCGTAGAAATGGTTATGCCGGTCA	55941
Qy	361	AsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLysPhe	380
Db	55940	ANTGTAAATATCAGTGGAACTTATTGCTCTGTGGCACTTGAAGATGGAACCTGTTTT	55881
Qy	381	AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleGlu	399
Db	55880	GCGATTGCTGAAGGTGGTAGAACAGTTGTTCAAGTGTGTTGAACAAAAATATTGAG	55824
RESULT 6			
LOCUS	BX571658	346792 bp	DNA linear BCT 08-SEP-2003
DEFINITION	Wolinella succinogenes, complete genome; segment 2/7.		
ACCESSION	BX571658 BX571656		
VERSION	BX571658.1 GI:34482500		
KEYWORDS	complete genome.		
SOURCE	Wolinella succinogenes		
ORGANISM	Wolinella succinogenes		
REFERENCE	1. Schuster, S.C., Baar, C., Eppinger, M., Raddatz, G., Simon, J.M., Lanz, C., Klimmek, O., Nandakumar, R., Gross, R., Rosinus, A., Keller, H., Jagtap, P., Linke, B., Meyer, F., and Lederer, H. Complete genome sequence and analysis of Wolinella succinogenes Unpublished		
JOURNAL	Schuster, S.C.		
REFERENCE	2. Direct Submission		
AUTHORS	Submitted (15-MAY-2003) Max-Planck Institut for Developmental Biology, Spenannstr. 35, 72076 Tuebingen, GERMANY		
JOURNAL	Location/Qualifiers		
FEATURES	1. 346792		
source	/organism="Wolinella succinogenes"		
gene	/mol_type="genomic DNA"		
CDS	/strain="DSMZ 1740"		
	/db_xref="taxon:844"		
	53..1297		
	/locus_tag="WS0352"		
	53..1297		
	/locus_tag="WS0352"		
	/codon_start=1		
	/transl_table=11		
	/product="PUTATIVE PROCESSING PROTEASE"		
	/protein_id="CAE09501.1"		
	/db_xref="GI:34482501"		
	/translation="MAAQLSEIEVKGVIPLIYEKSTLMPAVNIQIVTHAGSITDEI PLGLSARIFNEGTGTLGSEVFAKLEENRAISLSVSGNETLVFDMALKSEYLEGL RLTRLELRQPNFTKEALSQVKTDIKASLLRKESDFDYVAOVELHRRILPEGTPLVPSP GNRESIKITLKEVEEFKKRGVLRRAILVGGDVEEVEKKDLSLSVLPLGEGE LVRYEAGDLEVKSVIKNTQAYIIFGSPFVKNLREDSYKAKVAAPVLGSSGSGSRM MEELVRKRLAYSAYRISLKSQVNSYSLGTLKLENEQQQAIGVVRLEIEFLQKGIT EKELEGAQKPLLGSEPLRNETLSQRLSFAFSAYYRGLPLDPAKEELHKIEALTLEIN AYIKEHQELLQLSFSVVTPKE"		
	1297..3114		
	/gene="RECG"		
	/locus_tag="WS0353"		
	1297..3114		
	/gene="RECG"		
	/locus_tag="WS0353"		
	/EC_number="3.6.1.-"		
	/codon_start=1		
	/transl_table=11		
	/product="ATP-DEPENDENT DNA HELICASE EC 3.6.1."		
	/protein_id="CAE09502.1"		
	/db_xref="GI:34482502"		
	/translation="WEPSDLKLKSLKCKDALDFALLAPKGVDRRPSSELVSGAEGV VETIELSFQKQKRGARGLAWLERKERAELIIFHPKYYHHALPAPKRFYKAKVRR FGKLLIQPKQIAKIGILPKFTKVPFNQTLIELASRVLSKESLIALGIPALYASRL CELFFPDSDFLRRYQBERGIPKEHQEALKFAEIPHLRALYVKKRSFEALCRCONAPW		

gene
CDS
/gene="DNAB"
/locus_tag="WS0354"
3218..4639
/gene="DNAB"
/locus_tag="WS0354"
/EC_numbers="3.6.1.-"
/codon_start=1
/transl_table=11
/product="REPLICATIVE DNA HELICASE EC 3.6.1."
/protein_id="CAE09503.1"
/db_xref="GI:34482503"
BALXELRLDLPLOEEFIVKXETPERPIDRELLYLIGTNPISIEAVNIEKRSIK
RELRLANTIRYSQESLESEQLDHESELYKIQDNENRDFKAKEVTLATLEHI
KMKERGTILIGVDITGFKELNKMTHFGHGLDVIVAARPAMGKTFVLNMAQALDT
GKVAIFSEPAEQMLRMLSAKTSISLQNLRVNLQDEWEERLSRASDEMAKKPLF
VDDNSLITNOLRSKRLKRGHPVGMIIIDYLQMSGTGRDRHLEVSISGLKM
LARELEIPIIALSOLNRSLESRDERPMLSDLRSGATEQADAILFVYREAVYKMD
EKEKEQAKSGEVRNPFVNKEEERIIIGKQRNGPIGTVKLIFHKHCTRFVDFDK
EVEIVFEGKFDITNEAKIEMASI"
4686..5894
/gene="COMEC"
/locus_tag="WS0355"
4686..5894
/gene="COMEC"
/locus_tag="WS0355"
/codon_start=1
/transl_table=11
/product="DNA TRANSFER PROTEIN"
/protein_id="CAE09504.1"
/db_xref="GI:34482504"
/translation="MALLIVGLLSMGFEYHAFKKUREEPNKLQAVLLQYTKDEKGE
RIVKLQTEQGVQYVTHKEPLQDLTRYVSLTLPQKITFDLYKGFAPSYWIIIR
PERDFRGLRQIEFQHQKEALAGSLYRTFLADPLEKSLREGASRLGISHLLAISGPH
LGVLSGVLWGIWFPYRFLORRFFPYRNAFYDLGALTLLCMFGYLLLGSGSPSLRHH
AMGVWGYPLLHGRGKILSFSSLLAVGIMLVALFRLLFVSGVFFLSMGVVFLLFLHRLH
TAHLKQWNVFILLNCLYFWMLIPVHSIPETFPSPYQLFSPLLTMTLFTLYFLALFLHL
VGGMGMDWVFLGGLGGTESISVRATWFWAGYGIALLALLANRYLWATVGRGALCMGF
FFYAIYFLFG"
5947..6489
/locus_tag="WS0356"
5947..6489
/locus_tag="WS0356"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE09505.1"
/db_xref="GI:34482505"
/translation="MEDIFRISKGLLTILFPVILVWIFSVYKMLAGIPFYLRGMDT
SNLFATUIMFCSLIVFIYIGLVEKNRFLLLKFTLIVIERIPVWKGVYATVDKLIK
IFSGGDKENYLGVAIFYAGGSKLIGFTKEBEDGYWIFVPTTNPNTTGLLKIKEEV
ELSNMSISDGFKKIIISLVK"
6493..7356
/gene="UBIA"
/locus_tag="WS0357"
6493..7356
/gene="UBIA"
/locus_tag="WS0357"
/codon_start=1
/transl_table=11
/product="4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE"
/protein_id="CAE09506.1"
/db_xref="GI:34482506"
/translation="MNFSKVLQTLKFNELVMFQHTIFSLFFPIFIAMITTAAGGWFGWK
LLLLGLGAASARNFAFNAFNRWCDDRIDAKNPRATSPSVDRGVGGGKMWLFIILGNCA

LFVGVAYLINDLAFKLSPIIGILALYSVMKRFSPFLAHLILGLSLALAPIAGVIASVE
SVPLWSPLSGVLFWAGFDVLYSLODIEFDKNEGLHSIPSRGEKAILLSRIFHL
FTLFPFWLFIAPFLQIMLPAMAFVCWMLGYEHLVYSQDFRNPFAKFTVNGYLGIV
PLTFCIADRAL"
7353..7838
/locus_tag="WS0358"
7353..7838
/locus_tag="WS0358"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE09507.1"
/db_xref="GI:34482507"
/translation="MSDGRVVVRVPLSLEWERVEARGEFEREFLLSLDLEDDPVGOW
VMAKAKGEAKYSEDLTKLVLHQKIDTTLTKLQNESRYPALAFKAGVSEIGHV
VVMQESVFKSQELYYARINLPVFPQIVPPFFRGAGGGKASENPRQRSEGLHLYR
L"
7891..8391
/locus_tag="WS0360"
7891..8391
/locus_tag="WS0360"
/note="CONSERVED HYPOTHETICAL 19.7 KDA PROTEIN"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE09508.1"
/db_xref="GI:34482508"
/translation="WTWIMAGGGALFLLLLAYLYLKEMETNKLRLHFEKSIIEELNKQ
LYLQKKIKEDVESEFSSALSLSLRQEVKEAVYNNAAAGIYQSVRVEALWMBHRD
VDEKIVMLEERVKEMGYFTSPNGVDEARILSMFKDGSIDSIAKELRIGKGEVEFTL
KLANIE"
8491..9036
/locus_tag="WS0361"
8491..9036
/locus_tag="WS0361"
/note="CONSERVED HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE09509.1"
/db_xref="GI:34482509"
/translation="MSSKSPYFSSILFAALIVLSNFTVQVSYNDLYTYGALTYPTPTFL
LADVLSRYAKEVLKVVRIGIVLAPVPSLILASDWRIALASVGAFPIAQLDVLVYY
LKEKFKPLWLRNNGSTMLSQWDTMIFFHVAFLFVMPWPVIMALGDYAIKLLAL
LDTPFLYLLAIRVQKRLGIXR"
complement(9026..9352)
/locus_tag="WS0362"
complement(9026..9352)
/locus_tag="WS0362"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE09510.1"
/db_xref="GI:34482510"
/translation="MANSLYLPALLSDEQLGALKKEAKIKENPDTEIPLMQAITDL
ALNAGKALESHQIEALETYLRVLSLEKLHRLGIIKEYEQSPKEPFLPAKEITLTL
LTPPSA"
complement(9355..9858)
/gene="PPIA"
/locus_tag="WS0363"
complement(9355..9858)
/gene="PPIA"
/locus_tag="WS0363"
/EC_number="5.2.1.8"
/codon_start=1
/transl_table=11
/product="PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
/protein_id="CAE09511.1"

gene
CDS
/gene="DNAB"
/locus_tag="WS0354"
3218..4639
/gene="DNAB"
/locus_tag="WS0354"
/EC_numbers="3.6.1.-"
/codon_start=1
/transl_table=11
/product="REPLICATIVE DNA HELICASE EC 3.6.1."
/protein_id="CAE09503.1"
/db_xref="GI:34482503"
BALXELRLDLPLOEEFIVKXETPERPIDRELLYLIGTNPISIEAVNIEKRSIK
RELRLANTIRYSQESLESEQLDHESELYKIQDNENRDFKAKEVTLATLEHI
KMKERGTILIGVDITGFKELNKMTHFGHGLDVIVAARPAMGKTFVLNMAQALDT
GKVAIFSEPAEQMLRMLSAKTSISLQNLRVNLQDEWEERLSRASDEMAKKPLF
VDDNSLITNOLRSKRLKRGHPVGMIIIDYLQMSGTGRDRHLEVSISGLKM
LARELEIPIIALSOLNRSLESRDERPMLSDLRSGATEQADAILFVYREAVYKMD
EKEKEQAKSGEVRNPFVNKEEERIIIGKQRNGPIGTVKLIFHKHCTRFVDFDK
EVEIVFEGKFDITNEAKIEMASI"
4686..5894
/gene="COMEC"
/locus_tag="WS0355"
4686..5894
/gene="COMEC"
/locus_tag="WS0355"
/codon_start=1
/transl_table=11
/product="DNA TRANSFER PROTEIN"
/protein_id="CAE09504.1"
/db_xref="GI:34482504"
/translation="MALLIVGLLSMGFEYHAFKKUREEPNKLQAVLLQYTKDEKGE
RIVKLQTEQGVQYVTHKEPLQDLTRYVSLTLPQKITFDLYKGFAPSYWIIIR
PERDFRGLRQIEFQHQKEALAGSLYRTFLADPLEKSLREGASRLGISHLLAISGPH
LGVLSGVLWGIWFPYRFLORRFFPYRNAFYDLGALTLLCMFGYLLLGSGSPSLRHH
AMGVWGYPLLHGRGKILSFSSLLAVGIMLVALFRLLFVSGVFFLSMGVVFLLFLHRLH
TAHLKQWNVFILLNCLYFWMLIPVHSIPETFPSPYQLFSPLLTMTLFTLYFLALFLHL
VGGMGMDWVFLGGLGGTESISVRATWFWAGYGIALLALLANRYLWATVGRGALCMGF
FFYAIYFLFG"
5947..6489
/locus_tag="WS0356"
5947..6489
/locus_tag="WS0356"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE09505.1"
/db_xref="GI:34482505"
/translation="MEDIFRISKGLLTILFPVILVWIFSVYKMLAGIPFYLRGMDT
SNLFATUIMFCSLIVFIYIGLVEKNRFLLLKFTLIVIERIPVWKGVYATVDKLIK
IFSGGDKENYLGVAIFYAGGSKLIGFTKEBEDGYWIFVPTTNPNTTGLLKIKEEV
ELSNMSISDGFKKIIISLVK"
6493..7356
/gene="UBIA"
/locus_tag="WS0357"
6493..7356
/gene="UBIA"
/locus_tag="WS0357"
/codon_start=1
/transl_table=11
/product="4-HYDROXYBENZOATE OCTAPRENYLTRANSFERASE"
/protein_id="CAE09506.1"
/db_xref="GI:34482506"
/translation="MNFSKVLQTLKFNELVMFQHTIFSLFFPIFIAMITTAAGGWFGWK
LLLLGLGAASARNFAFNAFNRWCDDRIDAKNPRATSPSVDRGVGGGKMWLFIILGNCA

LFVGVAYLINDLAFKLSPIIGILALYSVMKRFSPFLAHLILGLSLALAPIAGVIASVE
SVPLWSPLSGVLFWAGFDVLYSLODIEFDKNEGLHSIPSRGEKAILLSRIFHL
FTLFPFWLFIAPFLQIMLPAMAFVCWMLGYEHLVYSQDFRNPFAKFTVNGYLGIV
PLTFCIADRAL"
7353..7838
/locus_tag="WS0358"
7353..7838
/locus_tag="WS0358"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE09507.1"
/db_xref="GI:34482507"
/translation="MSDGRVVVRVPLSLEWERVEARGEFEREFLLSLDLEDDPVGOW
VMAKAKGEAKYSEDLTKLVLHQKIDTTLTKLQNESRYPALAFKAGVSEIGHV
VVMQESVFKSQELYYARINLPVFPQIVPPFFRGAGGGKASENPRQRSEGLHLYR
L"
7891..8391
/locus_tag="WS0360"
7891..8391
/locus_tag="WS0360"
/note="CONSERVED HYPOTHETICAL 19.7 KDA PROTEIN"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE09508.1"
/db_xref="GI:34482508"
/translation="WTWIMAGGGALFLLLLAYLYLKEMETNKLRLHFEKSIIEELNKQ
LYLQKKIKEDVESEFSSALSLSLRQEVKEAVYNNAAAGIYQSVRVEALWMBHRD
VDEKIVMLEERVKEMGYFTSPNGVDEARILSMFKDGSIDSIAKELRIGKGEVEFTL
KLANIE"
8491..9036
/locus_tag="WS0361"
8491..9036
/locus_tag="WS0361"
/note="CONSERVED HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE09509.1"
/db_xref="GI:34482509"
/translation="MSSKSPYFSSILFAALIVLSNFTVQVSYNDLYTYGALTYPTPTFL
LADVLSRYAKEVLKVVRIGIVLAPVPSLILASDWRIALASVGAFPIAQLDVLVYY
LKEKFKPLWLRNNGSTMLSQWDTMIFFHVAFLFVMPWPVIMALGDYAIKLLAL
LDTPFLYLLAIRVQKRLGIXR"
complement(9026..9352)
/locus_tag="WS0362"
complement(9026..9352)
/locus_tag="WS0362"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE09510.1"
/db_xref="GI:34482510"
/translation="MANSLYLPALLSDEQLGALKKEAKIKENPDTEIPLMQAITDL
ALNAGKALESHQIEALETYLRVLSLEKLHRLGIIKEYEQSPKEPFLPAKEITLTL
LTPPSA"
complement(9355..9858)
/gene="PPIA"
/locus_tag="WS0363"
complement(9355..9858)
/gene="PPIA"
/locus_tag="WS0363"
/EC_number="5.2.1.8"
/codon_start=1
/transl_table=11
/product="PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
/protein_id="CAE09511.1"

Alignment Scores:
Pred. No.: 3.62e-113 Length: 346792
Score: 1862.00 Matches: 361
Percent Similarity: 94.96% Conservative: 16


```

/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB74180.1"
/db_xref="GI:6967818"
/db_xref="UniProt/TREMBL:Q9PIF7"
/translation="MEITDLPYLIIGIISGLGFGIGGMIIVPSMFALCASAHHA
IGISVLQMFYRAVEGSYNYKKNLNLDGIMIGFGGLIGASFSGMLKALSDVALTS
VFLAVSCFFIKYARIGIKENIVQNRSVWVKNVILFIAGATGIFAIISGIGGGLLIA
FIGLYFDGSKKVVLSLFFVIFASVSGIISFGNSGVSEIVHKGILVGIASWVG
PIKAIIEKMHISAKRKILLCVYALSILGTTSHSLNKLNF"
1009..1134
/gene="Cj0344"
/EC_number="4.1.3.27"
1009..1134
/gene="Cj0344"
/EC_number="4.1.3.27"
/translation="Cj0344", unknown, len: 41 aa; no Hp match"
/codon_start=1
/transl_table=11
/product="hypothetical protein Cj0344"
/protein_id="CAB74181.1"
/db_xref="GI:6967819"
/db_xref="UniProt/TREMBL:Q9PIF6"
/translation="MFQNIKYKDFIIFILNKLQNYLLIKINLDFKRFHKSLENF"
1299..2549
/gene="trpE"
1299..2549
/gene="trpE"
/EC_number="4.1.3.27"
/translation="Cj0345, trpE, possible anthranilate synthase
component I, len: 416 aa; similar in C-terminal half to
many e.g. TRPE CLOTM anthranilate synthase component I (EC
4.1.3.27) (494 aa), fasta scores; opt: 784 z-score: 1209.9
E(): 0, 47.0% identity in 281 aa overlap. 35.6% identity
in 365 aa overlap to Hpi282. Contains Pfam match to entry
PF00425 chorismate bind, chorismate binding enzyme, score
309.50, E-value 3.9e-89"
/codon_start=1
/transl_table=11
/product="putative anthranilate synthase component I"
/protein_id="CAB74182.1"
/db_xref="GI:6967820"
/db_xref="GOA:Q9PIF5"
/db_xref="UniProt/TREMBL:Q9PIF5"
/translation="MFVKVNPYVYQILEKSPENSYFADLTKEIIGIDCVLDNELS
FSEKAKYALUSKNKICDFAFGFVFSANVSLPEKIPLSKKNYDFPLFLFANAKA
YLIYKNSKMPFKGASKYFYLKDDIEPMKKQKNDPEILNSLEKNDKFKMCEKA
KEYLSGDIQVVLVSGQICIKHQVNAFYESLSALNPSYMFPPSYGVVLGSSPE
FLKIKKEIYVLAPTAGTNLENCDLLEKLSLSEKLSLSEKLSLSEKLSLSEKLSL
GTQTRVENLFSIINKKFMVHIVSEVYANMKEDASIFDVIEAVFPAGTISGAPKIRALE
ITSELEDGRTGYGAVGFLNPNEDITLAILRCFAFFQDKAYLASGAGIVLOSQK
EYAEICARRKALLVAFENLKKENQ"
1743..2522
/gene="trpE"
/translation="Cj0346, trpD (trpGD), probable anthranilate
synthase component II, len: 533 aa; (contains glutamine
amidotransferase and anthranilate
phosphoribosyltransferase) similar
to many e.g. TRPE_ECOLI
anthranilate synthase component II (EC 4.1.3.27) (530 aa),
fasta scores; opt: 866 z-score: 781.3 E(): 0, 32.6%
identity in 528 aa overlap. N-terminus has 34.6% identity
to Hpi281, C-terminus is 33.2% identity to Hpi280.
Contains PF00442 Glutamine amidotransferases class-I
active site and Pfam matches to entry PF00117 Gatase,
Glutamine amidotransferases class-I, score 252.00, E-value
8.4e-72 and to entry PF00591 Glycosyl transferase family,
transferase family, score 138.60, E-value 1.1e-37"
/codon_start=1
/transl_table=11
/product="anthranilate synthase component II"
/protein_id="CAB74183.1"
/db_xref="GI:6967821"
/db_xref="GOA:Q9PIF4"
/db_xref="UniProt/TREMBL:Q9PIF4"
/translation="WILLIDNVDSPFVNVKSMLEQLNDEILVRNDALISLSEIKQLN
PTHLILSPGKHPQSQCICLIEIFKARLNIPVLGICLGHQALALAFDSLVVKMQEPMHA
KNSLILKQRENELFSPNSFVSMRYHSLEVKQLSDELEILALDEKGVIMALGHKNLP
YVQVQHPFYSFYGLQFNFLKQDIKPKQKQENPLSFYLOKQSENHFLOQSDDFEQ
ICKIMSKDYELIQVALLIILITKESLNKLSLAFVQILRYSTQFSDSEMDICGT
GCDGFSINTVSTTSFAIILAGLVKAGNRAISSSGSTVDLALNLTPTNTLESVL
KQNNQGLSFLHAPFPHPLVGLKELIRSLRVTVENVLGLPMLKLVQLMGVNH
APVHLLLEVLNLRKALVNRGNDGMDLSICDESKIYELCEGILLEYSCIPQFG
FKRAPHSEIIGSSADNADKLDILSGRMQQAQKFDLVVLMAMFALYTANKASSPLVAK
DMILEAIYSGKVIYFKEYQAYAKA"
2546..3103
/gene="trpD"
/translation="Cj0347, trpF, probable
N-(5'-phosphoribosyl)anthranilate isomerase, len: 199 aa;
similar to many trpF e.g. TRPF LACCA
N-(5'-phosphoribosyl)anthranilate isomerase (199 aa) (EC
5.3.1.24), fasta scores; opt: 271 z-score: 352.4 E():
2.3e-12, 31.0% identity in 200 aa overlap, and to the trpF
domain of many trpC proteins e.g. TRPC SALTY
indole-3-glycerol phosphate synthase, fasta scores; opt:
307 z-score: 394.5 E(): 1e-14, 30.8% identity in 201 aa
overlap. 38.2% identity to C-terminus of Hpi279 (trpC)
Contains Pfam match to entry PF00697 PRAI,
N-(5'-phosphoribosyl)anthranilate (PRA) isomerase, score
102.30, E-value 9.5e-27"
/codon_start=1
/transl_table=11
/product="N-(5'-phosphoribosyl)anthranilate isomerase"
/protein_id="CAB74184.1"
/db_xref="GI:6967822"
/db_xref="GOA:Q9PIF3"
/db_xref="UniProt/Swiss-Prot:Q9PIF3"
/translation="WLKIKGIDKKNADKALFNIIDFFGLIFAKSPRRVLSLEQARN
LSAIPHEKDKVGVFVDENLEQLIRCEKAKLDGIQIYRTITKEEFILKQNVFWM
QVIVSENSLDLSEIFANLVLDAGILKGGNGISFDWLLIGSYTKDFILAGGIGLDN
VHKAVTKGAKILDKLSLEDEKGLDKINKIRKQILKELKK"
4143..4721
/gene="trpF"
/translation="Cj0348, trpB, probable tryptophan synthase beta
chain, len: 392 aa; highly similar to many e.g. TRPB THEME
tryptophan synthase beta chain (EC 4.2.1.20) (389 aa),
fasta scores; opt: 1476 z-score: 2082.2 E(): 0, 59.9%

```

identity in 384 aa overlap. 56.4% identity to H21278.

Contains F500168 Tryptophan synthase beta chain
Pyridoxal-phosphate attachment site, P500017
ATP/GTP-binding site motif A (P-loop), and pfam match to
entry PF00247 trp synthB, Tryptophan synthases, beta chain,
score 806.40, E-value 1e-238"
/codon_start=1

Alignment Scores:

Pred. No.: 3,34e-107 Length: 308601
Score: 1772.00 Matches: 336
Percent Similarity: 93.23% Conservative: 36
Best Local Similarity: 84.21% Mismatches: 27
Query Match: 86.74% Indels: 0
DB: 1 Gaps: 0

US-10-039-183a-4 (1-399) x CU11168X2 (1-308601)

```

QY      1 MetAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
DB      120309 ATGGCTAAAGAAATTTTCAGTAATAAGCCACACGTAATATTGGTACTATTGGTCAT 120368

QY      21 ValAspHisGlyLysThrThrLeuSerAlaLaiLeSerAlaValLeuSerLeuLysGly 40
DB      120369 GTTGACCATGTAAACACTACTTTACAGCTGCTATTTCGCTGTTCTTCTAGAGAGGT 120428

QY      41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
DB      120429 TTAGCAGAGCTAAAGATTATGATAATATCGATAATGCTCCAGAGAAAGAGCGTGGT 120488

QY      61 IleThrIleAlaThrSerHisIleGluThrGluThrGluAsnArgHisTyrAlaHisVal 80
DB      120489 ATTACTATTGCTACTTCTCATATTGAATATGAACAGACAATTCGTCACACAGCTT 120548

QY      81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGlnMetAsp 100
DB      120549 GACTGCCAGGTCACCGAGATTATGTTAAACATGATTACAGGTCTGCACAAATGAT 120608

QY      101 GlyAlaIleLeuValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
DB      120609 GGAGCGATCTGGTTGTTCTGCTCAGATGSCCTATGCCACAACATAGAGAGCACATT 120668

QY      121 LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMet 140
DB      120669 CTTCTTCTCGTCAAGTAGGCGTTCATATATTGTTGTTTATGAATAAAGCAGATATG 120728

QY      141 ValAspAspGlnGluLeuGluLeuValGluMetGluValArgGluLeuLeuSerAla 160
DB      120729 GTTGATGATGCTGAACCTTTAGAGTAGTTGAATGGAATGGAATATAGAGAAATTTATTAAGCTCT 120788

QY      161 TyrGluPheProGlyAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
DB      120789 TATGATTTCCAGCGGATGATACACTATTATTCTGTTCTGCTTTAAAGCTCTTGA 120848

QY      181 GluAlaLysAlaGlyAsnValGlyGluTrpGlyGluLysValLeuLysLeuMetAlaGlu 200
DB      120849 GAAGCTAAAGCTGGCAAGATGGTGAATGCTCAGCAAAAATATAGATCTTATGGCTGCA 120908

QY      201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
DB      120909 GTTGATGACTATATTCCAACTCCAACTCGTGATATCGTGAATGGAAGAGCTCTTGTATGCCAAT 120968

QY      221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
DB      120969 GAAGATGTTTCTCAATTTCCAGTCTGCTGCTGTTGTTTACAGGTAGTAATTTGAAAAGGT 121028

QY      241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThr 260
DB      121029 GTTGTAAAGTAGGTGATCTATCGAATCGTTGTTGTTATTAAGATACTCAAAACAACACT 121088

QY      261 ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal 280
DB      121089 GTACAGGTGTTGAAATGTTTCAGAAAAGAAATGGATCAAGGCGAAGCAGGAGATAACGTA 121148

```

```

QY      281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys 300
DB      121149 CGTGTCTTCTTCGTGGTACTAAAAAGAAAGAGTATATCCGTGGTATGTTCTTCTGCTAA 121208

QY      301 ProGlySerIleThrProHisLysLysPheGluGlyGluIleThrValLeuSerLysGlu 320
DB      121209 CCAATATCAATTAATCTCCACACTGACTTCGAGCTGAGTTATATCTTAATAAAGAT 121268

QY      321 GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArgThr 340
DB      121269 GAAGGTGGTAGACATACTCCATCTTTAAACAACATATAGACACACAGCTTTTATGTAAGAACA 121328

QY      341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
DB      121329 ACTGATGTTACAGGTTCGATTAATTAATGCTGATGCTGTTGAAATGCTTATGCCAGGTGAA 121388

QY      361 AsnValLysIleThrValGluLeuLeuSerProValAlaLeuGluLeuGlyThrLysPhe 380
DB      121389 AATGTGAGAAATTAATGTAAGCTTGATCGCTCAGTAGCATTGGAAGAAGGAACCTGTTTT 121448

QY      381 AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu 399
DB      121449 GCTATTCGTGAAGGTGGTAAACATGTTGCTTCAGGTGTTGTTCTTAAATATTATTA 121505

```

RESULT 10

```

AY305395          1185 bp      DNA      linear      SYN 26-SEP-2003
LOCUS             ML-meso gene, complete cds.
DEFINITION        AY305395
ACCESSION         AY305395
VERSION           AY305395.1 GI:32186877
KEYWORDS          synthetic construct
SOURCE            synthetic construct
ORGANISM          other sequences; artificial sequences.
REFERENCE         1 (bases 1 to 1185)
AUTHORS           Gaucher,E.A., Thomson,J.M., Burgan,M.F. and Benner,S.A.
TITLE             Inferring the palaeoenvironment of ancient bacteria on the basis of
                  resurrected proteins
JOURNAL           Nature 425 (6955), 285-288 (2003)
MEDLINE           2860410
PUBMED           13679914
REFERENCE         2 (bases 1 to 1185)
AUTHORS           Gaucher,E.A., Thomson,J.M., Burgan,M.F. and Benner,S.A.
TITLE             Direct Submission
JOURNAL           Submitted (22-MAY-2003) Chemistry, NASA Astrobiology Institute, PO
                  Box 117200, Gainesville, FL 32611-7200, USA
FEATURES          Location/Qualifiers
                   1..1185
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                   1..1185
                     /note="BLANKET-3"
                     /codon_start=1
                     /transl_table=11
                     /product="reconstructed ancestral elongation factor Tu
                     ML-meso"
                     /protein_id="AAP72171.1"
                     /db_xref="GI:32186878"

```

CDS

```

/translation="MAKEPERTPHVNIIGTIGHVDHGKTTLTAAITTVLAKKGIAEEA
RAYQIDNAPEEKERGITINISHVEYTEKRYAHVDCPGHADYVKNMIVTGAQMDGA
ILVVAATGMPQTRREHILLARQGVPIVVALNKCNDVDBDELELLEMEVREILSS
YDFGDDTPVIRNSALKALEGDEKWEKLELMDADVDEIPTPERDTPKPLMPEDV
FTITGRGVTVRVERGVLKGVDEIVIGIKETKTIVTIGEMFKRLKLDAGQNVG
LLLRGIKREDYVERGQVLAKPGSIPTHTKFEAEVYVLSKEGGRTHTFFNNRYPQYFR
TTDVTGVITLPEGTEMVMPGDNVENTVELIAPIMEGLRFAIREGGRTVGAGRVTIK"

```

ORIGIN

Alignment Scores: 1.8e-101 Length: 1185
Pred. No.:

```
Score: 1649.50 Matches: 315
Percent Similarity: 88.97% Conservative: 40
Best Local Similarity: 78.95% Mismatches: 39
Query Match: 80.74% Indels: 5
DB: 12 Gaps: 1

US-10-039-183A-4 (1-399) x AY305395 (1-1185)

Qy 1 MetAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
Db 1 ATGGCGAAGAAAGTTTGAACGTACAAAACCGCACGTTAACTTGGTACTATCGGCCAC 60
Qy 21 ValAspHisGlyLysThrThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLysGly 40
Db 61 GTTGACCATGTTAAACAACGCTGACCGCTGCAATCACTACCGTACTGGCTAAAAAGGC 120
Qy 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
Db 121 CTGGCTGAAGCTCGCGCATATGACCATCGATACGATACGCGCGGAGAAAGAAAGACGTGGT 180
Qy 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnAtcHisTyrAlaHisVal 80
Db 181 ATCCATCAACATTTCTCAGCTGAATACGAACCGAAACGTCATCGCACACGTA 240
Qy 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGluAlaAlaGlnMetAsp 100
Db 241 GACTGTCGGGCGACCGCATGTTTAAACATGATCACGGTGTGTCACAGATGGAC 300
Qy 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
Db 301 GCGCGGATCTGTTAGTTCAGCAACTGATGTCGCGATCGCGCAGACTCGTGAGCACATC 360
Qy 121 LeuLeuSerArgGlnValGlyValProHisIleValPheLeuAsnLysGlnAspMet 140
Db 361 CTGTGGCTCGTCAGTGGCGTTCGTCATCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 420
Qy 141 ValAspAspGlnGluLeuLeuGluLeuValGluMetGluValArgGluLeuSerAla 160
Db 421 GTTGATGACGAGAACTCTGGAACCTGTTGAAATGGAGTTCGTGAGCTTCTGTTCT 480
Qy 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 481 TAGCACTTCGATGGCGGACGACACTCCGGTTATCCGCTTATCCGCTTCTGCTCTGAAAGCACTGGAA 540
Qy 181 GluAlaLysAlaGlyAsnValGlyGluTyrGlyGluLysValLeuLysLeuMetAlaGlu 200
Db 541 GCGCGAGAA-----AAGTGGGTTGAGAAATCTCGAACTGATGGACGCT 585
Qy 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
Db 586 GTGGATGATATCATTTCCGACCCAGCGTGATACCGACGCGCTTCTGATGCCGATT 645
Qy 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
Db 646 GAAGACGATTACCATCACTACCGGTCGTGGTACGGTTGTTACGGTTCGTGTAGAACCGCGT 705
Qy 241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThr 260
Db 706 GTTCTGAAAGTTGGTGATGAAGTTGAAATCGTTGGTATCAAGAGACACTCAGAAGACACCC 765
Qy 261 ValThrGlyValGluMetPheArgLysGluLeuLysGlyGluAlaGlyAspAsnVal 280
Db 766 GTGACCGCATGTAATGTTCCGAAACTGTCGACGAGGAGCCGCGCTGGTGACACGTA 825
Qy 281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys 300
Db 826 GGTCTGCTGCTGCGTGGTATCAACGCTGAAGATGTCGAACGCTGGTCACTGGCTAAG 885
Qy 301 ProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGlu 320
Db 886 CCGGCGACGATCACGCGGCACACCAAGTTTGAAGCAGAAAGTGTACGTCTCTGCCAAGAA 945
Qy 321 GluGlyGlyArgHisThrProPheThrAenTyrArgProGlnPheTyrValArgThr 340
```

```
Db 946 GAAGTGGCCGCTACTCCGTTCTTCAATAACCTACCGTCGCGAGTTCTACTTCGCTACT 1005
Qy 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 1006 ACTGACGTGACTGGTGTGATTACTCTGCCGAAGCTACCGAGATGTTAATGCCGGCGAC 1065
Qy 361 AsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLysPhe 380
Db 1066 AACGTGGAATGACCGTTGAACCTGATCGCCCGGATCGCCATGGAAGAGGCTTCGTTTC 1125
Qy 381 AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleGlu 399
Db 1126 GCAATCCGTGAAGCGGTCGTACCGTTGGCGCGGCGCGTGAACCAAAATATCAAA 1182

RESULT 11
AY305397 1185 bp DNA linear SYN 26-SEP-2003
LOCUS Synthetic construct reconstructed ancestral elongation factor Tu
DEFINITION Alt-stem gene, complete cds.
ACCESSION AY305397
VERSION AY305397.1 GI:32186881
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1185)
AUTHORS Gaucher, E.A., Thomson, J.M., Borgan, M.F. and Benner, S.A.
TITLE Inferring the palaeoenvironment of ancient bacteria on the basis of
JOURNAL Nature 425 (6955), 285-288 (2003)
MEDLINE 22860410
PUBMED 13679914
REFERENCE 2 (bases 1 to 1185)
AUTHORS Gaucher, E.A., Thomson, J.M., Borgan, M.F. and Benner, S.A.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2003) Chemistry, NASA Astrobiology Institute, PO
Box 117200, Gainesville, FL 32611-7200, USA
FEATURES
source
1..1185
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
CDS
1..1185
/note="BLANKET-2"
/codon_start=1
/transl_table=11
/product="reconstructed ancestral elongation factor Tu
Alt-stem"
/protein_id="AAP72173.1"
/db_xref="GI:32186882"
/translation="MAKEPERTKPHVNI GTIGHVDHGKTTLTAAITKVLKGLAE
KAYDOIADNPEEKARGITINITHVEYTERKHAYHIDCPGHADYIKNMITGAQMDDGA
ILVVAATDGMPTQREHLVARQVPIYVFLNKCMDVDDLELELLELVEYMERLLSK
YDFGDDVPVIRGSKALGENDNEWKPILEMDADVNIIPDPERDVDPKPLMPTEDV
FSITGRGTVTVTRIERGVKPGDEIVILGELTKTITVTSVEMFKLDEGAGDNVG
CLLGLIEKEVERGQVLAKGSIPTHTKFEAQVVLKKEGGRHTPEFNRYRPOPYER
TTDTVTGIVELPEGVEMVMPGDNVENTVELIYPIALIEGLRPAIRREGRTVGAGVTKI
IE"
ORIGIN
Alignment Scores:
Pred. No.: 6,17e-101 Length: 1185
Score: 1641.50 Matches: 312
Percent Similarity: 87.47% Conservative: 37
Best Local Similarity: 78.20% Mismatches: 45
Query Match: 80.35% Indels: 5
Gaps: 1
DB: 12
US-10-039-183A-4 (1-399) x AY305397 (1-1185)
Qy 1 MetAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
```



```

Db 1 ATGGCGAAGAAAGTTTGAACGTCACAAAACCGCACGTAACATTTGGTACTATCGGCCAC 60
Qy 21 ValAspHisGlyLysThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLysGly 40
Db 61 GTTGACCAACGCAAAACCAACGCTGACCGCTGCAATCACTAAAGTACTAGCTTGAAGGC 120
Qy 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
Db 121 CTGGCTGAAGCTAAAGCATATGACAGATCATACGCTCCAGATCCAGAAAGAAAGACGCGGT 180
Qy 61 IleThrIleAlaThrSerHisIleGluThrGluAsnArgHisTyrAlaHisVal 80
Db 181 ATCAACATCAATACCACTGCAATGCAACGAAACGAAACGTCACCTACGACACAT 240
Qy 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGlnMetAsp 100
Db 241 GACTGTCCGGGCGACGACATATATTAACATGATCAGCGGTGCTGCACAGATGAC 300
Qy 101 GlyAlaIleLeuValSerAlaAlaAspGlyProMetProGlnThrArgGluHisLe 120
Db 301 GCGCGATCTCTGGTAGTTGCGACCACTGATGCTCGATCGCGACGACTCGTGAACATGTG 360
Qy 121 LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMet 140
Db 361 CTGCTGGCTCTCAGTAGGCGGTCCTGACATCGTCGTTCTGTAACAAATGTGACATG 420
Qy 141 ValAspGlnGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
Db 421 GTTGATGACGAAGAACTGCTGAACTGTTGAAATGGAAGTTCGTGACCTCTCTCTAA 480
Qy 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 481 TACGACTTTCCAGGTGATGACGTGCGGCTTATCTGCTGTTCTGCGCTGAAAGCACTGAA 540
Qy 181 GluAlaLysAlaGlyAsnValGlyValGluThrGlyLysValValLeuLysLeuMetAlaGlu 200
Db 541 -----GGCGACAATGAATGGTATTAACCGGCTCTGGAAGTATGATGACCGCT 585
Qy 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
Db 586 GTGGACAACATACATCCCGGACCCAGAGCGTGATGTTGACAGCGCTTCTCTGATCGCGATT 645
Qy 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
Db 646 GAAGACGCTATTACGACTTACTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705
Qy 241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThr 260
Db 706 GTTATTAAACCGGGTGATGAAGTTGAAATCGTGGGTCTGGAAGACACTAAAGAAACCCACC 765
Qy 261 ValThrGlyValGluMetPheArgLysGluLeuLysGlyGluAlaGlyAspAsnVal 280
Db 766 GTGACTTCTGTTGAATGTTCCGCAACCTGTCGACGAAGTCAAGCTGGTGTGTAACGTA 825
Qy 281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys 300
Db 826 GGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
Qy 301 ProGlySerIleThrProHisLysPheGluGlyGluIleTyrValLeuSerLysGlu 320
Db 886 CCGGGTAGCATCAGCCACACACCAAGTTTGAAGCACAAGTGTACGCTCTGAAAGAAAGAA 945
Qy 321 GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArgThr 340
Db 946 GAAGGTGGCGTCACTCTCCGTTCTTCAATACCTACCGTCCGCGATCTCTCTCTCTCTCTCT 1005
Qy 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 1006 ACTGACGTGACTGGTATTGTTGAATCGCGGAAGGTGGAGATGGTAAATGCCGGCGAC 1065
Qy 361 AsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLysPhe 380
Db 1066 AACGTGGAATGACCGGTGAATCTATCTATCTCCGATTGCAATGGAAGAGGTCTTCTGTTTT 1125

```

```

Qy 381 AlaIleArgGluGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu 399
Db 1126 GCAATTCGTGAAGGTGGTCTGACTGTAGTGGCGGCGTGTGAACCAAAATTATCGAA 1182

RESULT 12
AY305396 1185 bp DNA linear SYN 26-SEP-2003
LOCUS ML-stem gene, complete cds.
DEFINITION Synthetic construct reconstructed ancestral elongation factor Tu
ACCESSION AY305396
VERSION AY305396.1 GI:32186879
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthet sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1185)
AUTHORS Gaucher, E.A., Thomson, J.M., Burgan, M.F. and Benner, S.A.
TITLE Inferring the paleoenvironment of ancient bacteria on the basis of
JOURNAL Nature 425 (6955), 285-288 (2003)
MEDLINE 22860410
PUBMED 13679914
REFERENCE 2 (bases 1 to 1185)
AUTHORS Gaucher, E.A., Thomson, J.M., Burgan, M.F. and Benner, S.A.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2003) Chemistry, NASA Astrobiology Institute, PO
Box 117200, Gainesville, FL 32611-7200, USA
FEATURES
source
1..1185
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
CDS
1..1185
/notes="BLANKET-1"
/codon_start=1
/transl_table=11
/product="reconstructed ancestral elongation factor Tu
ML-stem"
/protein_id="AAP72172.1"
/db_xref="GI:32186880"
/translation="MAKEPERTKPHVNVGTIGHVDHGTTLTAITTVLALKGLAE
KAYDQIDKAPEEKARGITINIAHVEYETERKHYAHIDCPGHADYIKNMITGAQMDGA
ILVVAATDGPMPQTRHILLARQGVYIVVINKVMDVDDPPELLELVEMVRDLISK
YFPGDVPVIRSGALKALEAPQKYEKILELLDAVDIEYIPTPRDVKPKPLMPEDV
PSTGRGTVTVTGRIERGVIKPGDEIVLGLSETKTKTTSVEMFRKLDEGLAGDNVG
VLLRGIDKEVERGOVLAKSGITPHTKFAQVVLKKEGGRHTPFENGVRPOFYFR
TTDVTGIVELPEGVMPVMPGDNVENTVELIYPIAMEGLRPAIREGRTVAGVVTKI
IE"
ORIGIN
Alignment Scores:
Pred. No.: 1,15e-99 Length: 1185
Score: 1622.50 Matches: 308
Percent Similarity: 86.97% Conservativeness: 39
Best Local Similarity: 77.19% Mismatches: 47
Query Match: 79.42% Indels: 5
DB: Gaps: 1
US-10-039-183A-4 (1-399) x AY305396 (1-1185)
Qy 1 MetAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
Db 1 ATGGCGAAGAAAGTTGAACGTCACAAAACCGCACGTAACATTTGGTACTATCGGCCAC 60
Qy 21 ValAspHisGlyLysThrLeuSerAlaAlaIleSerAlaValLeuSerLeuLysGly 40
Db 61 GTTGACCAACGCAAAACCAACGCTGACCGCTGCAATCACTAAAGTACTAGCTTGAAGGC 120
Qy 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
Db 121 CTGGCTGAAGCTAAAGCATATGACAGATCATACGCTCCAGATCCAGAAAGAAAGACGCGGT 180

```

```
Qy 61 IletThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
Db 181 ATCCACATCAACATTGCACACATCGANTACGAACCGAAACGACGTCATCCACACATT 240
Qy 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGlnMetAsp 100
Db 241 GACTGTCCGGGGCACGACATATATAAAACATGATCACGGGTGTCACAGATGGAC 300
Qy 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
Db 301 GCGCGGATCCTGTAGTTGCACCACTGATGTGTCGATCGCCACACATCGTGAACATATA 360
Qy 121 LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMet 140
Db 361 CTGCTGGCTGTCAGTAGGCGTTCGTACATCGTGTGTTTATTAAACAAAGTTGACATG 420
Qy 141 ValAspAspGlnGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla 160
Db 421 GTTGATGACCCGGAACCTGCGAACTGGTTGAAATGGAAGTTTCGTGATCTTCTGCTAAA 480
Qy 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 481 TACGAATTTCCAGGTGATGAAGTCGCGTATTTCGTGTTCTGCGCTGAAAGCACTGGA 540
Qy 181 GluAlaLysAlaGlyAsnValGlyGluTyrPglyGluLysValLeuLysLeuMetAlaGlu 200
Db 541 GCTCCGCA-----AAATGGTATGAATAAATCTCGAAGCTGTGGACGCT 585
Qy 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
Db 586 GTGGACGAATACATCCGACCCAGAGCGTGATGTTGACAAAGCGTCTCTGATGCGGATT 645
Qy 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
Db 646 GAAGACGTATTACGATATTACTGCTGCTGCTACGTTGTTACGGGTGCTGTAACGAGTGT 705
Qy 241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThr 260
Db 706 GTTATTAAACCGGGTATGAAGTTGAATCTGGGTCTGAGCGAGACTAAATAAACCAACC 765
Qy 261 ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal 280
Db 766 GTGACTTCTGTTGAAATGTTCCGCAAACTGCTGGACGAAGGTATTGCTGGTGATAACGTA 825
Qy 281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys 300
Db 826 GGTGTTCTGCTGCGTGGTATTGATAAAGAAGAAGTCGAACGCTGGTCAGGTACTGCGTAAG 885
Qy 301 ProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGlu 320
Db 886 CCGGTAGCATCACCCACACCAAGTTTAAAGCACAGTGTACGTCTCTGNAATAAGAA 945
Qy 321 GluGlyGlyArgHisThrProPheThrAsnTyrArgProGlnPheTyrValArgThr 340
Db 946 GAAGGTGCGGTGATCTCCGTTCTTCAATGGCTACCGTCCGACGTCTACTTCGCTACT 1005
Qy 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 1006 ACTGACGTGACTGGTATTGTTGAACCTGCGGAAGGTGTGGAGATGGTAAATGCCGGCGAC 1065
Qy 361 AsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLysPhe 380
Db 1066 AACGTGGAATAGACCGTGGAACTGATCTATCCGATTGCCATGGGAAGAGGTTCTCGTTTT 1125
Qy 381 AlalLeArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu 399
Db 1126 GCAATTCGTGAAGTGGTCTGACTCTAGTGTGCGGCGTGTGTAACCAAAATTATCGAA 1182
RESULT 13
AL646073/c 204050 bp DNA linear BCT 02-SEP-2002
LOCUS Ralstonia solanacearum GMI1000 chromosome, complete sequence;
DEFINITION segment 17/19.
```

```
ACCESSION AL646073 AL646052
VERSION AL646073.1 GI:17429991
KEYWORDS
SOURCE Ralstonia solanacearum
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE
1 Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Ariat,M., Billault,A., Bottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
Sigulier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Nature 415 (6871), 497-502 (2002)
JOURNAL
MEDLINE 21681879
PUBMED 11823852
REFERENCE 2 (bases 1 to 204050)
AUTHORS Boucher C.A.
DIRECT SUBMISSION
SUBMITTED (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
COMMENT http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
location/Qualifiers
1..204050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
complement(115..1278)
/gene="dgt"
/feature(115..1278)
complement(115..1278)
/feature(115..1278)
/EC_number="3.1.5.1"
/function="small molecule metabolism; central intermediary
metabolism; nucleotide hydrolysis"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE DEOXYGUANOSINE TRIPHOSPHATE
TRIPHOSPHOHYDROLASE PROTEIN"
/db_xref="GI:17429992"
/db_xref="GOA:Q8XV63"
/db_xref="UniProt/Swiss-Pro:Q8XV63"
/transl_table="MNDRLDLPAPFDLEGLAPVAHAHAOSRGVHAEPSTSTSTEF
ORDRDRIHCTAFRLRYKTVFVNHGDLFRTRLTHSLEVAQIARSRLNDEL
VEALSADHLGHTPFHGAGQALNDKMPYGGFHNLSQSLLVVLRLERYGFGDGLNL
TFETREGILKHCNSNAATLDGLRRFLGEGQPSLEAQLANLADAVYNNHDIDGLR
SGLTTLQLEGVPLWARRAEAAFPVCGGRMTNIRIINALIYDLIAGTAAAI
TGAAPQSIDVRAAPLVAFPSEPMRDEARVLKFLFNLRYHLYVMRAAKARIIED
LFRVLEDPRLLPQYQARDPADQPRWIAHYIAGTWDRYAIKEHRIIFAVEVL"
complement(1298..2404)
/gene="arob"
/note="synonyms: RS01328, RSC2969"
complement(1298..2404)
/gene="arob"
```


QY	61	lleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal	80
Db	72869	ATCAGATCAACACCGGCACATCAGTACGAAACGCGCAACCGCACTACGACACGTC	72810
QY	81	AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGlyAlaAlaGlnMetAsp	100
Db	72809	GACTGCCGGCGCACCGGACTACGTCAAGAACATGATCACCGGTGGCGCCAGATGGAC	72750
QY	101	GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle	120
Db	72749	GGCGCATCTGCTGCTCGCGCGCTGACGCGCGGATCGCCCAACCGCTGAGCACATC	72690
QY	121	LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMet	140
Db	72689	CTGCTGGCGCGCGCGGCTGCGGTGCGTACATCATCTCTTCTCGAACAAAGTGCACATG	72630
QY	141	ValAspAspGlnGluLeuLeuGluValGluMetGluValArgGluLeuLeuSerAla	160
Db	72629	GTGGAGGACGCTGAGCTGCTGGAACCTGGTCTGAAATGGAAGTGGCGAGCTGCTCCAAG	72570
QY	161	TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu	180
Db	72569	TACGACTTCCCGGGCGGACACACCCCGCATCATCAAGGGTTTCGCCCAACCGCTGGAA	72510
QY	181	GluAlaLysAlaGlyAsnValGlyGluTrpGlyGlu---LysValLeuLysLeuMetAla	199
Db	72509	-----GGCGCAAGGGCGAGTGGCGGAAGTGGCCATCAACCTGGCCGAC	72462
QY	200	GluValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetPro	219
Db	72461	GCACGTGGACTCGTACATCCGACGCGGAGCGCGTGTGCGGCACGTTCTCTGATGCG	72402
QY	220	ValGluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArg	239
Db	72401	GTGGAAGACGTTCTCGATCTCGGTCTCGGCACCGTGTGACCGCGCGTATCGAGCGC	72342
QY	240	GlyValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThr	259
Db	72341	GGCATCATCAAGTTCGGCGGAATAATCGTTCGGCATCAAGCCACCGCAAGAAC	72282
QY	260	ThrValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsn	279
Db	72281	ACCTGCACCGGCTGGAAATGTTCCGCAAGCTGCTGGACCAAGTTCAGGCTGGCGAAC	72222
QY	280	ValGlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCys	299
Db	72221	GTGCGTATCTGCTGCGCGGACCAAGCGCGAAGACGTCGACGCGCGCAAGTCTGTGC	72162
QY	300	LysProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLys	319
Db	72161	AAGCGGGTTCGATCAAGCGGCACACGCACTTCACCGCGGAGGTCTACATCTGTCGAAG	72102
QY	320	GluGluGlyGlyArgHisThrProPheThrAsnTyrArgProGlnPheTyrValArg	339
Db	72101	GACGAAGTGTGTCACACCCCGTCTTCAACAATATCGCCCGAGGTCTATATTTCGCG	72042
QY	340	ThrThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGly	359
Db	72041	ACGACCGAGTGCACCGGCTCGATCGAGTCCGCGAAGGCAAGAAATGTCATGCCCGCG	71982
QY	360	AspAsnValLysIleThrValGluLeuLeuSerProValAlaLeuLeuLysGlyThrLys	379
Db	71981	GACACGTTGATCATCCGCTCAAGCTGATCGCCCGATCGCCATCGAAGAGGCTCGCG	71922
QY	380	PheAlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu	399
Db	71921	TTCGCCATCCGTGAAGGGCTGTAACCTCGCGCGCGGCGTCTGCCCAAGATCATCGAG	71862
RESULT 14			
LOCUS	AE005800	7934 bp	DNA linear BCT 12-JUN-2002
DEFINITION	Caulobacter crescentus CB15 section 126 of 359 of the complete		

genome.	AE005800	AE005673
VERSION	AE005800.1	GI:13422568
KEYWORDS	Caulobacter crescentus CB15	
SOURCE	Caulobacter crescentus CB15	
ORGANISM	Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales; Caulobacteraceae; Caulobacter.	
REFERENCE	1 (bases 1 to 7934)	
AUTHORS	Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and Fraser, C.M.	
TITLE	Complete genome sequence of Caulobacter crescentus	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)	
MEDLINE	21173698	
PUBMED	11259647	
REFERENCE	2 (bases 1 to 7934)	
AUTHORS	Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	
FEATURES	Location/Qualifiers	
source	1..7934	
	/organism="Caulobacter crescentus CB15"	
	/mol_type="genomic DNA"	
	/strain="CB15"	
	/db_xref="taxon:190650"	
	119..202	
	/gene="tRNA-Tyr-1"	
	119..202	
	/gene="tRNA-Tyr-1"	
	/product="tRNA-Tyr"	
	313..386	
	/gene="tRNA-Gly-3"	
	313..386	
	/gene="tRNA-Gly-3"	
	/product="tRNA-Gly"	
	441..1631	
	/gene="CC1240"	
	441..1631	
	/gene="CC1240"	
	/note="identified by match to TIGR protein family HMM TIGR00485"	
	/codon_start=1	
	/transl_table=11	
	/product="translation elongation factor EF-Tu"	
	/protein_id="AAK23221.1"	
	/db_xref="GI:13422569"	
	/translation="MAKEPERTKPHCNIGTIGHVDHGKTTLTAAITWTAKSGGATA KKYDEIDAAPEEKARGITINTAHVEYETANHYAVDCPGHADYVKNMITGAQMDGA ILVSDAAGPMPQTRHILLARQGVPAIVVPMKYNVDDEELLELEVEMSVRELSS YQFGDDIPITKGSAALAAVEGRDPOIGEEKILELMSVDAITPQEPVDMFPLMPE DVFISGRGTGVTVRGVIGVEEIVIGIRVPQKTTCTGTFEMFRKLKLDQGOAGDN VGLLRGTGRDVERGOVLCKPFGSITPTTKFEVAEAYILTKEEGRHTFPFTNRPQFY FRTDVTGIIKLRGVEIMPGDNDALDELITPIAMEKLRFAIREGRTVAGVA KIVE"	
	1724..2353	
	/gene="CC1241"	
	1724..2353	
	/gene="CC1241"	
	/note="identified by Glimmer2; putative"	

Db 1152 ATCGGAAGTCGGTGAAGAAGTCGAATCGTCGGCATCCGTCGGTTTCAGAAAGCAGACC 1211
Qy 261 ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspSerVal 280
Db 1212 TCACGGGGCGTTGAATGTTCCGAAGTCGTGGCAAGGTCGAAGGTCGAAGCGCGCACACGTTG 1271
Qy 281 GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys 300
Db 1272 GCGGTGCTGCTCGCGGGACCAAGCGTGAAGACGTCGAGCGCGGCAGGTCGTGTGCAAG 1331
Qy 301 ProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGlu 320
Db 1332 CCGGTTTCGATCACCCCGCACACCAAGTTCTGTCGGCGAAGCCTACATCTGACCAAGGAA 1391
Qy 321 GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArgThr 340
Db 1392 GAAGCGCGCGCTCACACCCCGTCTTCCACCAACTACCGTCGCGAGTCTTACTTCGCGACG 1451
Qy 341 ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp 360
Db 1452 ACGGACGTGACGGGATCATCAAGCTGCGGAAGCGGTGGAATGATCATGCCGGCGGAC 1511
Qy 361 AsnValLysIleThrValGluLeuLysSerProValAlaLeuGluLeuGlyThrLysPhe 380
Db 1512 AACCGCGAGCTGACGTCGAGCTGATCACCCGATCGCATGGAAGAGAGCTGCGCTTC 1571
Qy 381 AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu 399
Db 1572 GCCATCGTGAAGCGCGCGCACCGCTCGCGCGCGGTGGTTCGGAAGATCGTTGAA 1628

RESULT 15
LOCUS AE005984 10212 bp DNA linear BCT 12-JUN-2002
DEFINITION Caulobacter crescentus CB15 section 310 of 359 of the complete genome.
ACCESSION AE005984 AE005673
VERSION AE005984.1 GI:13424876
KEYWORDS
SOURCE Caulobacter crescentus CB15
ORGANISM Caulobacter crescentus CB15
Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
Caulobacteraceae; Caulobacter.

REFERENCE 1 (bases 1 to 10212)
AUTHORS Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and Fraser, C.M.
TITLE Complete genome sequence of Caulobacter crescentus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
MEDLINE 21173698
PUBMED 11259647

REFERENCE 2 (bases 1 to 10212)
AUTHORS Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers

1..10212
/organism="Caulobacter crescentus CB15"
/mol_type="genomic DNA"
/strains="CB15"
/db_xref="taxon:190650"

gene
CDS

complement (142..1332)
/gene="CC3199"
complement (142..1332)
/gene="CC3199"
/note="identified by match to TIGR protein family HMM TIGR00485"
/codon_start=1
/transl_table=11
/product="translation elongation factor EF-Tu"
/protein_id="AAK25161.1"
/db_xref="GI:13424877"

/translation="MAKEFERTPHNCNIGTIGHVDHGKTTLTAAITMTLAKSGGATA
KKYDEIDAPEKARGITINTAHVEYETANRHYAHVDCPHADYVKNMITGAOAMDGA
ILVSAADGPMPTREHILLAROVGPALVVMKNDVDDSELELLEMEVEFLLSS
YQFQDDIPITKGSALAAVEGRDQIGEEKILELMASVDAYLPQEPVDMFLMPVE
DVFSISGRGTVTVRVERGIVKVEEVEIVGRPVQKTTCTGVMFRKLDDQGGADN
VGILLRTKREDVERGQVLCKPQSTTPTHTKFAVAYILTKBEGGRHTTFFNTYRQFY
PRTTDTVTGIIKLREGVEMIPGDNALDELVELITPIAMEEKLRFIREGGRTVGAGVA
KIVB"

gene
CDS

complement (1452..3530)
/gene="CC3200"
complement (1452..3530)
/gene="CC3200"
/note="identified by match to TIGR protein family HMM TIGR00484"
/codon_start=1
/transl_table=11
/product="translation elongation factor G"
/protein_id="AAK25162.1"
/db_xref="GI:13424878"

/translation="MPRTKIEDYRNFIMAHIDAGKTTTTERILYYTGKSHKIGEVH
DGAATMDMEOEORGITITSAATAFWMDKRLNIDTPGHVDPTIEVERSLRVLDDGA
VTVLDDGNAGVEPQETWRQADKYKRVIVFVNKMDKIGADFDSVESIRDLGAKAV
PIQPIGESNLKGLVDLRWKA VVWMDNDGUGASTRDEIPADLMDKAVEARAYLVEN
AVELEDDAMEAYLGGEEPSIETIKKIRKAVLTGAFYKILCGSAFKNKGVLPLDADV
DYLSPVDIPPTKIDFKTEETTRKASDEEPLSVAFKIMDDPFPVGLTFCRIYSGK
METGMSLLNSTRDKRERVRLMHSNNREDIKEYAGDIVALAGLKEKTRTCDTLCDP
LKSPVILERMEFPAPVTEIAVEPKSKADOEKLVALOKLAEDPSFTVTDPESGOTI
LKNGELHLDIKIDILKTYKVEANIGAPOVAYRESLGRKVDIDYTHKOTGGTQFA
RWLITFEPGEGSFVFESALVGVAKEYIPGVQKGLSEKVDGSLLAGFLDLDFKAT
LTDQKLDVDSVLAFTASRAFKLEKRGAPKLEPIMKEVVTPEEYLGSLVIGDL
NSRRGIIQGGDMRGATVVAAYVPLANNFGYVNTLRGMSQGRAQFSWYDYHVPVQH
VADSVIKKYA"

gene
CDS

complement (3645..4118)
/gene="CC3201"
complement (3645..4118)
/gene="CC3201"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="ribosomal protein S7"
/protein_id="AAK25163.1"
/db_xref="GI:13424879"
/translation="MSRRRAEKQVLPDPKFGDLVVTKFMNYVMEGKKAVAENIY
RWLITAKRKDQGLPFTFHSALDNVAPAEVRSRRVGATQVPEVRRPDRRALAI
RMLVTAARKGENTMEKLAGELLDASNNRGRTAVKKREDTHKMAEANRAFSHYRW"

gene
CDS

complement (4132..4503)
/gene="CC3202"
complement (4132..4503)
/gene="CC3202"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="ribosomal protein S12"
/protein_id="AAK25164.1"
/db_xref="GI:13424880"
/translation="MPTINQIRKPRSPKPVNRKVPALKGCPQRRGVCTRVYTTTPPK
PNSALRKVAKVRLTTTGTEAVCYIPGEGHNLQHSVLTIRGVRKDLPGVRVHYHRLVL
DTQGVKDRKQRSLYGAKRPK"

gene
CDS

4838..5707
/gene="CC3203"
4838..5707
/gene="CC3203"

```
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25165.1"
/db_xref="GI:1342488"
/translation="MNGWSRLPWLRTGTFPGNAGRAPSGRGVQVSVMEGSEVERRR
ATTIAGSLHLLGALLAWLALPPPLLETFABDALPTMTLELQRLVLEGLVGRKLLIL
TATPASPFLVTRAPVPAPAPTPSLEVPASALARPVAIRPAPLPGDSAGDLTALR
GSATCCVNADAVGLNRRERCDEQFGAARAKGDPPLSGMDAGKRRALDAQAAOEAAYR
RYDAPMGPGVDHRSKHDPGTKEIPFVLGAQQDGLGRERKSVADEMRRANDLESRA
RYLELKAQDARR"
5756..6832
/gene="CC3204"
5756..6832
/gene="CC3204"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="hydrolase, alpha/beta hydrolase fold family"
/protein_id="AAK25166.1"
/db_xref="GI:1342488"
/translation="WIRALLAALSAASITVPDAARAAPWPTTEGLVFKDVTFKS
GERLSEARMRTTAGTTPHNAQGEIDNAVLLHGTGGSKNPLSPLEDELFGQPL
DLAKYVIMPONIHGSSKPSDGLRMAPPRIYDDIMIALQHLRLVLEGLVGRKLLIL
GTSQGMHAFVWGQTPGFARLAPFACNAVPLVGRNMRMRKMANDAIRADPAWKEGN
YTTPMAGRLTLLIILAGANPLAQQAQYPTREATKALDQAFNARIGTIDANDALY
YIDASRTYDPSGLEKIVPVLWNVSADDFINPPELGLAEPLAKRMPKARFVLIPAST
ETRGGHTAAKFWKADLAKLLAQ"
complement(6911..7468)
/gene="CC3205"
complement(6911..7468)
/note="CC3205"
/note="similar to SP:PI6921 PID:147801 PID:396321
GB:U00096 PID:1790414; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="transcription antitermination protein NusG"
/protein_id="AAK25167.1"
/db_xref="GI:1342488"
/translation="MSTETASNPNHKWTIVHAYNFEXKVAESTREQAKNQLEBNFS
EILPTEDVBIIRGRKVAERKFFPVLVKNLSDSEAYHLIKNTPKVTGFLSGSK
PMPSVEKEVRIIGTIEGVAAKPAVLVEYGVENRVVDGPFASFNGSVYVDSEKAR
LRVTSIFGRATPVELEYNQVEKIA"
complement(7479..7730)
/gene="CC3206"
complement(7479..7730)
/note="CC3206"
/note="identified by match to PFAM protein family HMM
PF00584"
/codon_start=1
/transl_table=11
/product="preprotein translocase, SecE subunit"
/protein_id="AAK25168.1"
/db_xref="GI:1342488"
/translation="MAPAGATATAAAPAPKPKRTNPLOFFREVRAEARKITWTTRKET
WITSVMVIMVLASLPRTAVDGGIGPAINQLLKLATAG"
complement(7895..7970)
/gene="trNA-Trp-1"
complement(7895..7970)
/gene="trNA-Trp-1"
/product="trNA-Trp"
complement(8401..8817)
/gene="CC3207"
complement(8401..8817)
/note="CC3207"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25169.1"
/db_xref="GI:1342488"
```

```
/translation="MRVWFTGAAAAASLVAFGLACSDSGAKSKVAQAOQVIVEATPV
TASGCVRPVENTNCLVWVRGGGYDSTASPADLSKGVASLQGMKMGKNTQCGRQ
LTDVRWSYLGICQCGTSVPASSATTTKDDTKTKVDKAG"
9330..10127
/gene="CC3208"
9330..10127
/gene="CC3208"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="ferredoxin-NADP reductase"
/protein_id="AAK25170.1"
/db_xref="GI:1342488"
7.58e-98 Length: 10212
Pred. No.: Matches: 310
Score: 1609.50
Percent Similarity: 86.97% Conservative: 37
Best Local Similarity: 77.69% Mismatches: 49
Query Match: Indels: 3
DB: Gaps: 2
US-10-039-183A-4 (1-399) x AE005984 (1-10212)
QY 1 MetalAlaLysGluLysPheAsnArgThrLysProHisValAsnIleGlyThrIleGlyHis 20
Db 1332 ATGCCCAAGGAAAGTTTCGAACGTAAGCCGCACTCAACATCGGCACCATTCGTTCAC 1273
QY 21 ValAspHisGlyLysThrLeuSerAlaIleSerAlaValLeuSerLeuLysGly 40
Db 1272 GTTGACCATGCAAGACGACGCTGACCGCCGCGATCAGATGACCTGGCGAAGTCGGGC 1213
QY 41 LeuAlaGluMetLysAspTyrAspAsnIleAspAsnAlaProGluGluLysGluArgGly 60
Db 1212 GCGCGACGCGCAAGAGTACGACGATATGACGCGCGCGCGAAGAAAGCCCGCGCG 1153
QY 61 IleThrIleAlaThrSerHisIleGluTyrGluThrGluAsnArgHisTyrAlaHisVal 80
Db 1152 ATCAGATCAACACCGCGCAGCTGGAATATGAGACGCGCAACCGCTACATCGCCACGTC 1093
QY 81 AspCysProGlyHisAlaAspTyrValLysAsnMetIleThrGluValAlaGlnMetAsp 100
Db 1092 GACTGCCCGCGCCAGCGCGACTACGTAGAAACATGATCAGCGCGCGCGCGAGATGAC 1033
QY 101 GlyAlaIleLeuValValSerAlaAlaAspGlyProMetProGlnThrArgGluHisIle 120
Db 1032 GCGCGATCTCTGTGTGTTTCGCGCGCTGACGCGCGATGCGCGACCGCGAGCACATC 973
QY 121 LeuLeuSerArgGlnValGlyValProHisIleValValPheLeuAsnLysGlnAspMet 140
Db 972 CTGTCGCGCGCTCAGTCGCGCTGCGCGCTCTGCTGCTGCTCATGAACAAGTCGACATG 913
QY 141 ValAspAspGlnGluLeuLeuValGluMetGluValArgGluLeuLeuSerAla 160
Db 912 GTCGACGACGAGAGAGCTCTGAGAGCTCTGTAAGATGGAAGTTCGCGAGCTGCTGAGACAGC 853
QY 161 TyrGluPheProGlyAspAspThrProIleValAlaGlySerAlaLeuArgAlaLeuGlu 180
Db 852 TACCAAGTTCGCGCGAGAGCATTCGATACCAAGGTTTCGGCTCTGCGCGCGCTT--- 796
QY 181 GluAlaLysAlaGlyAsnValGlyGluTrpGlyGluLysValLeuLysLeuMetAlaGlu 200
Db 795 GAAGTCTGTCGCGCGCAATCGCGAA-----GAGAAGATCTCTGAGCTGATGCTTCG 742
QY 201 ValAspAlaTyrIleProThrProGluArgAspThrGluLysThrPheLeuMetProVal 220
Db 741 GTCGACGCTATATCCCGACGCGCGAAGCCCGCGGACATCGCGCTTCTGATGCGCGTC 682
QY 221 GluAspValPheSerIleAlaGlyArgGlyThrValValThrGlyArgIleGluArgGly 240
Db 681 GAAGACGTGTTCTCGATCTCGGTCGCGGTACGTCGTCCAGCGTCGCGGTGAGCGCGCGC 622
QY 241 ValValLysValGlyAspGluValGluIleValGlyIleArgProThrGlnLysThrThr 260
```

Db	621	ATCGTGAAGGTCGGTGAAGAAGTCGAAATCGTCGCATCCCGTCCGGTTTCAGAAGACGACC	562
Qy	261	ValThrGlyValGluMetPheArgLysGluLeuGluLysGlyGluAlaGlyAspAsnVal	280
Db	561	TGCACGGCGGTGAATGTTCCGCAAGTGTCTGGACCAAGGTCAAGCCGGCGACACGTCG	502
Qy	281	GlyValLeuLeuArgGlyThrLysLysGluGluValGluArgGlyMetValLeuCysLys	300
Db	501	GGCGTGTCTGTCGGCGGCACCAAGCGTGAAGACGTCGAGCGCGCAGGTGTGTGCAAG	442
Qy	301	ProGlySerIleThrProHisLysLysPheGluGlyGluIleTyrValLeuSerLysGlu	320
Db	441	CCGGGTTTCGATCACCCCGCACACCAAGTTCGTGGCGCGAAGCCTACATCTGCACCAAGGAA	382
Qy	321	GluGlyGlyArgHisThrProPhePheThrAsnTyrArgProGlnPheTyrValArgThr	340
Db	381	GAAGCGCGCGTCACACCCCGTCTTTCACCAATACCGTCGCGAGTCTTACTTCGCGCAG	322
Qy	341	ThrAspValThrGlySerIleThrLeuProGluGlyValGluMetValMetProGlyAsp	360
Db	321	ACGACGTGACGGGCATCATCAAGCTGGCGGAGCGGTGGAATGATCATGCCGCGCGCAC	262
Qy	361	AsnValLysIleThrValGluLeuIleSerProValAlaLeuGluLeuGlyThrLysPhe	380
Db	261	AACGCCGAGCTGGACGTCGAGCTGATCACCCCGCATCGGCATGGAAGAGAAGCTGCGCTTC	202
Qy	381	AlaIleArgGluGlyGlyArgThrValGlyAlaGlyValValSerAsnIleIleGlu	399
Db	201	GCCATCCGTGAAGCGGCGGCACCGCTCGGCGCGCGTGTGGTGGCAAGATCTTCGAA	145

Search completed: April 24, 2005, 09:20:28
Job time : 6447.59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:31:54 ; Search time 1642.13 Seconds
(without alignments)
718.576 Million cell updates/sec

Title: US-10-039-183A-17
Perfect score: 31
Sequence: 1 gctaattattcaataattgtctcaac 31

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_ges1:
9: gb_ges2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23.6	76.1	958	8	CC266736 CH261-90B
C 2	23.2	74.8	550	1	AA516958 Vh85d10.1
C 3	22.6	72.9	418	8	BH612485 SALK_0328
C 4	22.2	71.6	751	9	AG394257 Mus_muscu
C 5	22	71.0	1092	9	AG088415 Pan_trogl
C 6	21.8	70.3	589	9	BX207173 Danio rer
C 7	21.6	69.7	416	1	AL381307
C 8	21.6	69.7	434	1	AL381308
C 9	21.6	69.7	524	9	CR325923
C 10	21.6	69.7	557	9	TA68E10P
C 11	21.6	69.7	567	9	TA68F10P
C 12	21.6	69.7	580	8	AQ491708
C 13	21.6	69.7	720	8	BH433207
C 14	21.6	69.7	782	8	BZ503289
C 15	21.6	69.7	832	8	BZ820968
C 16	21.6	69.7	885	8	BZ721736
C 17	21.6	69.7	948	9	CG110457
C 18	21.6	69.7	990	9	CL010460
C 19	21.4	69.0	399	9	DR39K4T
C 20	21.4	69.0	448	8	AZ161299
C 21	21.4	69.0	677	7	CO578276
C 22	21.4	69.0	691	7	CF114243
C 23	21.4	69.0	699	8	AZ138814
C 24	21.4	69.0	816	2	BF065214

25	21.4	69.0	826	2	BE559176
C 26	21.2	68.4	223	2	BF880225
27	21.2	68.4	489	6	CD812340
28	21.2	68.4	489	6	CD814287
29	21.2	68.4	489	6	CD815821
C 30	21.2	68.4	734	6	CB619111
C 31	21.2	68.4	1039	8	CC239056
C 32	21.2	68.4	1202	9	AG277970
C 33	21	67.7	231	8	CC027685
C 34	21	67.7	402	6	CB170034
35	21	67.7	416	4	B1208331
36	21	67.7	422	4	B1208557
C 37	21	67.7	437	9	CE024196
C 38	21	67.7	460	2	BF602438
C 39	21	67.7	510	2	BF602439
C 40	21	67.7	553	8	BZ276072
41	21	67.7	649	5	BM048970
42	21	67.7	666	8	BZ758001
43	21	67.7	671	5	BP131560
44	21	67.7	678	7	CNS14609
45	21	67.7	688	9	AG151595

ALIGNMENTS

RESULT 1 CC266736 958 bp DNA linear GSS 13-MAY-2003
LOCUS CH261-90B21.Sp6.1 CH261 Gallus gallus genomic clone CH261-90B21,
DEFINITION genomic survey sequence.
ACCESSION CC266736
VERSION CC266736.1 GI:30616030
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 958)
AUTHORS Warren, W., Graves, T., Mardis, E., and Wilson, R.
TITLE Gallus Gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 7
High quality sequence stop: 648.
Location/Qualifiers
1..958
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-90B21"
/sex="female"
/cell_line="CH261"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

ORIGIN
Query Match 76.1%; Score 23.6; DB 8; Length 958;
Best Local Similarity 86.7%; Pred. No. 1.9e+02;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 OCTAATATTATTCAATATATTGCTCAAA 30

Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@tc.riken.jp

PRIMERS
 Sequencing : TJ

LIBRARY

Vector : pBac3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1..751

/organism="Mus musculus molossinus"

/mol_type="genomic DNA"

/sub_species="molossinus"

/db_xref="taxon:57486"

/clone="MSMG01-213C16.TJ"

/sex="male"

/tissue type="mixture of kidney and spleen"

/clone_lib="MSMG01 Mouse Male BAC Library"

FEATURES

source

ORIGIN

Query Match 71.6%; Score 22.2; DB 9; Length 751;
 Best Local Similarity 88.9%; Pred. No. 6e+02;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 TAATATTATTCATAATATTTGCTCACA 29

|||||

Db 656 TAATATTATTCATAATATACTCTCA 682

RESULT 5

AG088415

LOCUS

DEFINITION Pan troglodytes DNA, clone: PTB-087F20.F, genomic survey sequence.

ACCESSION AG088415

VERSION AG088415.1

KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

PRIMERS

LIBRARY

Sequencing: -21M13

Vector

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1..1092

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="PTB-087F20.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

FEATURES

source

ORIGIN

Query Match 71.0%; Score 22; DB 9; Length 1092;
 Best Local Similarity 83.3%; Pred. No. 7.1e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CTATATTTCATAATATTTGCTCACAAC 31

|||||

Db 297 CTAATATAATAATAAATTTCTCACTAC 326

RESULT 6

EX207173/c

LOCUS

DEFINITION Danio rerio genomic clone DKEY-234E10, genomic survey sequence.

ACCESSION EX207173

VERSION EX207173.1

KEYWORDS GSS.

SOURCE Danio rerio (zebrafish)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

PEATUES

source

1..589

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="DKEY-234E10"

/tissue type="Testis"

/note="Vector pIndigoBAC-536"

Location/Qualifiers

589

Query Match 70.3%; Score 21.8; DB 9; Length 589;

Best Local Similarity 92.0%; Pred. No. 8.4e+02;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 ATATTATTCATAATATTTGCTCACA 29

|||||

Db 548 ATATTATTCATAATATTTACTCAA 524

RESULT 7

AL381307/c

LOCUS

DEFINITION MtBC019B10F1 MtBC Medicago truncatula cDNA clone MtBC019B1 T3, mRNA

sequence.

ACCESSION AL381307

VERSION AL381307.1

KEYWORDS EST.

SOURCE Medicago truncatula (barrel medic)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Genoscope

1 (bases 1 to 416)

Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,

Niebel,A., Carreau,V., Chatagnier,O., Kahn,D.,

Gianinazzi-Pearson,V. and Gamas,E.

Medicago truncatula ESTs from endomycorrhizal roots

Unpublished (2000)

Contact: Genoscope

Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email : secref@genoscope.cns.fr
Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
Mt-est@toulouse.inra.fr Website :
<http://sequence.toulouse.inra.fr/Mttruncatula.html>).

FEATURES

source

1. .416
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBC019B1"
/tissue_type="arbuscular mycorrhiza"
/dev_stage="harvested 3 weeks post inoculation with Glomus intraradices"
/clone_lib="MtBC"
/notes="Vector: pBluescript pSK; Site1: EcoRI; Site 2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Epcosso soil : 2/3 calcined terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LP48). The plants were watered every day and twice a week with a modified nutrient Long Ashton solution without phosphate but with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from poly(A)⁺ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Séquençage (Genoscope, Evry, France). Note : EST may be of fungal origin."

ORIGIN

Query Match	69.7%	Score 21.6;	DB 1;	Length 416;
Best Local Similarity	85.7%	Pred.No. 9.9e+02;		
Matches 24;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	3	TAATATTATTCAATAATATTGCTCACAA	30	
Db	322	TAATAATTATTCAATAATTTTTCACAAA	295	

RESULT 8	
AL381308/c	
LOCUS	434 bp mRNA
DEFINITION	AL381308 MCBC019B10R1 MtBC Medicago truncatula cDNA clone MtBC019B1 T7, mRNA sequence.
ACCESSION	AL381308
VERSION	AL381308.1 GI:9681059
KEYWORDS	EST.
SOURCE	Medicago truncatula (barrel medic)
ORGANISM	Medicago truncatula

REFERENCE	Medicago.
AUTHORS	1 (bases 1 to 434) Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.
TITLE	Medicago truncatula ESTs from endomycorrhizal roots
JOURNAL	Unpublished (2000)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage 2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE Email: segresf@genoscope.cns.fr Web : www.genoscope.cns.fr

Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : <http://sequence.toulouse.inra.fr/Mtruncatula.html>).

FEATURES

source

1. 434
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MCBC019B1"
/tissue_type="arbuscular mycorrhiza"
/dev_stages="harvested 3 weeks post inoculation with Glomus intraradices"
/clone_lib="McBC"
/note="Vector: pBluescript pSK; Site 1: EcoRI; Site 2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Eपोissees soil : 2/3 calcined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LPAG). The plants were watered every day and twice a week with a modified nutrient Long Ashton solution without phosphate but with a high level of nitrates. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using EXAssit helper phage and propagated in SOUR cells. Clone ordering and sequencing was performed by the Centre National de Sequençage (Genoscope, Evry, France). Note : EST may be of fungal origin."

ORIGIN

```

Query Match      69.7%; Score 21.6; DB 1; Length 434;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATAATAATTGCTCACAA 30
    |||||
Db 334 TAAATATTTCATAATAATTTTCTCAAAA 307
    |||||

```

RESULT	9
LOCUS	CR325923
DEFINITION	Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago truncatula, genomic survey sequence.
ACCESSION	CR325923
VERSION	CR325923.1 GI:44872067
KEYWORDS	GSS.
SOURCE	Medicago truncatula (barrel medic)
	linear DNA 524 bp
	GSS 01-MAR-2004

REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 524)
Genoscope.
Genoscope.
Direct Submission
Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
BP 131 91006 EVRY cedex - FRANCE (E-mail : segr@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Medicago.

FEATURES

source

```
1. .524
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="Jemalong A17"
/db_xref="taxon:3880"
/clone_lib="MTE1"
```

/note=Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
; Debelle F. and Chalhoub B.-Genoscope sequence ID :
mte1-53L24RM1"

ORIGIN

Query Match 69.7%; Score 21.6; DB 9; Length 524;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30

Db 218 TAATAATCTTCAATAATATTGCTCTCAA 245

RESULT 10

TA68E10P/c

LOCUS

DEFINITION T. brucei sheared genomic DNA clone 58e10, forward sequence, GSS 13-DEC-2000

ACCESSION AL457511

VERSION AL457511.1 GI:11858737

KEYWORDS GSS.

SOURCE Trypanosoma brucei

ORGANISM

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 557)

REFERENCE

AUTHORS

Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,

Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,

Melville, S.E., Rajandream, M.A. and Barrell, B.G.

Direct Submission

JOURNAL

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

COMMENT

Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source

1..557

/organism="Trypanosoma brucei"

/mol_type="genomic DNA"

/strain="TREU927"

/db_xref="taxon:5691"

/clone="68e10"

ORIGIN

Query Match 69.7%; Score 21.6; DB 9; Length 557;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30

Db 434 TAATATTATTGATATTGATAACAA 407

RESULT 11

TA68F10P/c

LOCUS

DEFINITION

T. brucei sheared genomic DNA clone 68f10, forward sequence, GSS 13-DEC-2000

ACCESSION AL457517

VERSION AL457517.1 GI:11858743

KEYWORDS GSS.

SOURCE ORGANISM

Trypanosoma brucei
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE

AUTHORS

1 (bases 1 to 567)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE

JOURNAL

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

COMMENT

Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source

1..567

/organism="Trypanosoma brucei"

/mol_type="genomic DNA"

/strain="TREU927"

/db_xref="taxon:5691"

/clone="68f10"

ORIGIN

Query Match 69.7%; Score 21.6; DB 9; Length 567;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30

Db 434 TAATATTATTGATATTGATAACAA 407

RESULT 12

AQ491708

LOCUS

DEFINITION

RPci-11-271F14.TJ RPci-11 Homo sapiens genomic clone

RPci-11-271F14, genomic survey sequence.

ACCESSION AQ491708

VERSION AQ491708.1 GI:4677582

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 580)

Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and

Venter, J.C.

Use of BAC End Sequences from Library RPci-11 for Sequence-Ready

Map Building

JOURNAL

COMMENT

Unpublished (1997)

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbs@tigr.org

Clones are derived from the human BAC library RPci-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: SP6
Class: BAC ends.

FEATURES

source
Location/Qualifiers
1..580
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7603813"
/db_xref="taxon:9606"
/clone="RPCI-11-271F14"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBac3.6; Site 1: EcoRI; Site 2: EcoRI;
RPC111 Human Male BAC Library"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 580;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATAATATTGCTCAAA 30

Db 284 TAATATTATTCATAAAATTTCTCAGAA 311

RESULT 13

BH433207/c
LOCUS
DEFINITION BH433207 720 bp DNA linear GSS 12-DEC-2001
BOGSW15TF BOGS Brassica oleracea genomic clone BOGSW15, genomic
survey sequence.

ACCESSION BH433207

VERSION BH433207.1 GI:17618928

KEYWORDS GSS.

SOURCE

ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 720)

AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other_GSSs: BOGSW15TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TP

Class: sheared ends.

FEATURES

source
Location/Qualifiers
1..720
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO100DH3"
/db_xref="taxon:3712"
/clone="BOGSW15"
/clone_lib="BOGS"
/note="Vector: pBOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pBOS1 using BstXI linkers"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 720;
Best Local Similarity 85.7%; Pred. No. 1e+03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATAATATTGCTCAAA 30

Db 318 TAATATTCTTCATTATATAGTCAAA 291

RESULT 14

BZ503289/c
LOCUS

DEFINITION

BZ503289

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 782)

AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other_GSSs: BONGE68TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..782

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO100DH3"

/db_xref="taxon:3712"

/clone="BONGE68"

/clone_lib="BO 1.6 2 KB tot"

/note="Vector: pBOS1; Site 1: BstXI; 1.6-2 kb sheared

total DNA inserted into pBOS1 using BstXI linkers"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 782;

Best Local Similarity 85.7%; Pred. No. 1e+03;

Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATAATATTGCTCAAA 30

Db 416 TAATATTCTTCATTATATAGTCAAA 389

RESULT 15

BZ820968

LOCUS

DEFINITION

BZ820968

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 832)

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and

Bennetzen, J.

TITLE Maize Genomics Consortium

JOURNAL Unpublished (2003)

COMMENT Other_GSSs: PUF6046TD

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

BZ503289 782 bp DNA linear GSS 16-DEC-2002
BONGE68TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONGE68,
genomic survey sequence.

BZ503289

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 782)

AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other_GSSs: BONGE68TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..782

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO100DH3"

/db_xref="taxon:3712"

/clone="BONGE68"

/clone_lib="BO 1.6 2 KB tot"

/note="Vector: pBOS1; Site 1: BstXI; 1.6-2 kb sheared

total DNA inserted into pBOS1 using BstXI linkers"

Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES

Source

Location/Qualifiers

1..832
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTaj14H19"
/clone_lib="ZM_0.6_1.0_KB"
/note="Vector: PCR1-TOPO, site_1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 832;
Best Local Similarity 85.7%; Pred. No. 1e+03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 4 AATATTATTCATATATATTGCTCACAC 31
||||| ||||| ||||| ||||| |||||
Db 770 AATATTATGAATATATAGCTACAC 797

Search completed: April 24, 2005, 05:28:25
Job time : 1643.13 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 17:27:10 ; Search time 804.69 Seconds
(without alignments)
1866.696 Million cell updates/sec

Title: US-10-039-183A-17

Perfect score: 31

Sequence: 1 gctaattattcaataattgtctcaac 31

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.on.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pc.*

10: gb.ro.*

11: gb.scs.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	31	6	BD082353
2	31	100.0	31	6	BD082355
C 3	31	100.0	1448	6	BD082347
C 4	31	100.0	2825	6	BD009865
5	31	100.0	10532	1	AE000626
6	31	100.0	15292	1	AE001541
C 7	29.4	94.8	968	6	BD061867
C 8	29.4	94.8	970	6	BD061998
9	25.4	81.9	30	6	AX793134
10	23.2	74.8	184688	2	AC102439
C 11	23.2	74.8	121995	10	AC107741
12	22.8	73.5	120794	2	AC017240
C 13	22.8	73.5	166863	3	AC009741
C 14	22.8	73.5	225655	3	AE003695
C 15	22.6	72.9	110000	1	EX571856_06
16	22.6	72.9	269050	1	AL935255
C 17	22.2	71.6	175477	5	AL954814
18	22.2	71.6	189773	2	CR376780
19	22.2	71.6	217717	2	CR293523

20	22	71.0	2532	5	AF081861
C 21	22	71.0	42627	3	AY190944
22	22	71.0	104995	9	AC136189
C 23	22	71.0	110000	1	EX571857_05
C 24	22	71.0	127922	9	AC134998
C 25	22	71.0	142250	10	AC132329
C 26	22	71.0	166974	2	AC046177
27	22	71.0	197208	10	AC122826
C 28	22	71.0	293050	1	AP003131
C 29	22	71.0	325350	1	AP004823
C 30	22	71.0	343590	1	AP003359
C 31	22	71.0	347572	2	AC107303
32	21.8	70.3	157216	5	EX897661
33	21.8	70.3	159610	5	EX088559
34	21.8	70.3	160487	5	EX005126
C 35	21.8	70.3	162785	5	EX547938
C 36	21.8	70.3	166133	9	AC092361
37	21.8	70.3	168189	2	EX470264
38	21.8	70.3	170524	2	AC138921
39	21.8	70.3	175762	2	EX957304
40	21.8	70.3	178408	2	CR450755
C 41	21.8	70.3	184142	9	AC138828
C 42	21.8	70.3	189043	2	AC139804
C 43	21.8	70.3	189471	2	AC139498
C 44	21.8	70.3	190095	2	AC139806
C 45	21.8	70.3	195579	2	AC139810

ALIGNMENTS

BD082353 31 bp DNA linear PAT 27-AUG-2002
76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
corresponding polynucleotide molecules.

ACCESSION BD082353.1 GI:22627963

VERSION JP 2001523954-A/60.

KEYWORDS Mastadenovirus

SOURCE Mastadenovirus

ORGANISM Mastadenovirus

REFERENCE 1 (bases 1 to 31)

AUTHORS Kleanthous, H., Lissolo, L., Tomb, J.F., Miller, C. and Garawi, A.A.

TITLE 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
corresponding polynucleotide molecules

JOURNAL Patent: JP 2001523954-A 60 27-NOV-2001;

MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET

VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,

HUMAN GENOME SCIENCES INC

COMMENT PN JP 2001523954-A/60

PD 27-NOV-2001

PF 31-MAR-1998 JP 1998541962

PI HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI

MILLER,

PI AMAL AL GARAWI

PC A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,

PC A61K39/40,

PC A61K39/40, G01N33/569

CC G01N33/54, G01N33/569

CC Strandedness: Single;

CC Topology: Linear;

FEATURES Key Location/Qualifiers.

source

1..31

/organism="Mastadenovirus"

/mol_type="genomic DNA"

/db_xref="taxon:10509"

ORIGIN

Query Match 100.0%; Score 31; DB 6; Length 31;

Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
    |||||
Db 1 GCTAATATTATTCATAATATTGCTCACAAC 31

RESULT 2
BD082355
LOCUS          31 bp DNA linear PAT 27-AUG-2002
DEFINITION    76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
               corresponding polynucleotide molecules.
ACCESSION     BD082355
VERSION       BD082355.1 GI:22627965
KEYWORDS      JP 2001523954-A/62.
SOURCE        Mastadenovirus
ORGANISM      Mastadenovirus
               Viruses; dsDNA viruses, no RNA stage; Adenoviridae.
REFERENCE     1 (bases 1 to 31)
AUTHORS       Kleanthous, H., Lissolo, L., Tomb, J.F., Miller, C. and Garawi, A.A.
TITLE         76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
               corresponding polynucleotide molecules
JOURNAL       Patent: JP 2001523954-A 62 27-NOV-2001;
               MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET
               VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
               HUMAN GENOME SCIENCES INC
COMMENT       PN JP 2001523954-A/62
               PD 27-NOV-2001
               PF 31-MAR-1998 JP 1998541962
               PI HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI
               MILLER,
               AMAL AL GARAWI
               PC A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,
               A61K39/40,
               PC G01N33/54, G01N33/569
               CC Strandedness: Single;
               CC Topology: Linear;
               FH Key Location/Qualifiers.

FEATURES             source
   source            1..1448
                     /organism="Mastadenovirus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10509"

ORIGIN
Query Match      100.0%; Score 31; DB 6; Length 1448;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
    |||||
Db 1324 GCTAATATTATTCATAATATTGCTCACAAC 1294

RESULT 4
BD009865/c
LOCUS          2825 bp DNA linear PAT 31-JAN-2002
DEFINITION    Proteins, in particular membrane proteins, of Helicobacter pylori,
               their preparation and use.
ACCESSION     BD009865
VERSION       BD009865.1 GI:18638238
KEYWORDS      JP 2001502886-A/6.
SOURCE        Helicobacter pylori
ORGANISM      Helicobacter pylori
               Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
               Helicobacteraceae; Helicobacter.
REFERENCE     1 (bases 1 to 2825)
AUTHORS       Knapp, B., Hundt, E. and Schmidt, K.H.
TITLE         Proteins, in particular membrane proteins, of Helicobacter pylori,
               their preparation and use
JOURNAL       Patent: JP 2001502886-A 6 06-MAR-2001;
               CHIRON BEHRING GMBH & CO
COMMENT       OS Helicobacter pylori
               PN JP 2001502886-A/6
               PD 06-MAR-2001
               PF 25-JUL-1997 JP 1998508651
               PR 26-JUL-1996 DE 196 30 390.7
               PI BERNHARD KNAPP, ERIKA HUNDT, KARL HEINZ SCHMIDT PC
               C12N15/31, C07K14/205, C07K16/12, G01N33/53, A61K31/70, A61K39/106, PC
               A61K39/395
               CC Strandedness: Single;
               CC Topology: Linear;
               FH Key Location/Qualifiers
               FT source 1..2825
               FT /organism="Helicobacter pylori"
               FT /mol_type="genomic DNA"
               FT /db_xref="taxon:210"

FEATURES             source
   source            1..2825
                     /organism="Helicobacter pylori"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:210"

ORIGIN
Query Match      100.0%; Score 31; DB 6; Length 2825;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
    |||||
Db 2097 GCTAATATTATTCATAATATTGCTCACAAC 2067

```



```

HEIKLKs"
/gene complement(3548.. .4747)
/cds /gene="HP1205"
/complement(3548.. .4747)
/note="similar to SP:P42482 PID:587590 percent identity:
89.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="translation elongation factor EF-Tu (tufB)"
/protein_id="AAD08250.1"
/db_xref="GI:2314366"
/translation="MAKEFNRTKPHVNIPTIGHVDHGGKTTLSAISAIVLSLUGLAEM
KDYNDIDNAPEKRGITTIATSHIYETENRHYAVDCPGHADYVKNMITGAQMDGA
ILVVSADGPMPTREHILLRSQVGVPHIVFLANKQDMDVDDQELLELMEMEVELLSA
YEFDDTPIVAGSALRALEAKAGNKGWKEKVLKMAEVDAYIPTPERDEKTFILM
PVDVFSIAGTGTVTGIRGVGVKGVDEIVGIRQTKTVTGTVGVNFKELKGEA
GDNVGLLRGTTKKEVERGMVLCKPGSITPHKKPEGEIYVLSKEGGRTPTFTNRP
QFYVRTDVTGTSITLPEGVEMVMPGDNVKITVELISVALELGTKFAIREGGRTVAG
VVSNIIE"
complement(4877.. .4948)
/gene="CRNA-Thr-1"
/complement(4877.. .4948)
/gene="CRNA-Thr-1"
/product="CRNA-Thr"
/note="codon recognized: ACC"
complement(4982.. .5055)
/gene="CRNA-Gly-2"
/complement(4982.. .5055)
/gene="CRNA-Gly-2"
/product="CRNA-Gly"
/note="codon recognized: GGA"
complement(5074.. .5155)
/gene="CRNA-Tyr-1"
/complement(5074.. .5155)
/gene="CRNA-Tyr-1"
/product="CRNA-Tyr"
/note="codon recognized: UAC"
complement(5186.. .5258)
/gene="CRNA-Thr-2"
/complement(5186.. .5258)
/gene="CRNA-Thr-2"
/product="CRNA-Thr"
/note="codon recognized: ACA"
5448.. .7184
/gene="HP1206"
/448.. .7184
/gene="HP1206"
/note="similar to SP:P22638 PID:142020 percent identity:
26.18; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="multidrug resistance protein (heta)"
/protein_id="AAD08251.1"
/db_xref="GI:2314367"
/translation="MAKKKHISTLKYLFLRSLKQIYMLITFKKMWFFLLVMAVFS
Query Match 100.0%; Score 31; DB 1; Length 10532;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTAATATATTCATATATTTGCTCAACAAC 31
Db 3541 GCTAATATATTCATATATTTGCTCAACAAC 3571
RESULT 6
LOCUS AE001541 15292 bp DNA linear BCT 20-JAN-1999
DEFINITION Helicobacter pylori, strain J99 section 102 of 132 of the complete
genome.
ACCESSION AE001541 AE001439
VERSION AE001541.1 GI:4155724

```

```

KEYWORDS Helicobacter pylori J99
SOURCE Helicobacter pylori J99
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
Helicobacteraceae; Helicobacter.
REFERENCE 1 (bases 1 to 15292)
AUTHORS Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C.,
Smith,D.R., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G.,
Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,
Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,
Vovis,G.F. and Trust,T.J.
TITLE Genomic-sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori
JOURNAL Nature 397 (6715), 176-180 (1999)
MEDLINE 99120557
PUBMED 9923682
2 (bases 1 to 15292)
AUTHORS King,B.L., Alm,R.A. and Trust,T.J.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
Street, Cambridge, MA 02139, USA
COMMENT Address all correspondence to: hp@arch.us.astro.com or Richard
A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and
Diane E. Taylor are with the University of Alberta Department of
Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G
2H7 and the Canadian Bacterial Diseases Network. All other authors
are with Astra Research Center Boston, 128 Sidney Street,
Cambridge, MA, 02139. Putative identifications, sequence
alignments, and name and sequence search capability are available
at ARCS's World Wide Web site. (URL:
http://www.astro-boston.com/hp/ylori).
FEATURES
source
1..15292
/organism="Helicobacter pylori J99"
/mol_type="genomic DNA"
/strain="J99"
/db_xref="taxon:85963"
/complement(257.. .415)
/gene="rpmG"
/note="synonym: jhp1127"
/complement(257.. .415)
/gene="rpmG"
/note="similar to H. pylori 26695 gene HP1204"
/codon_start=1
/transl_table=11
/product="50S RIBOSOMAL PROTEIN L33"
/protein_id="AAD06710.1"
/db_xref="GI:4155730"
/translation="MKWKIGLKCSDCEDINVTYKNAKNTKLEKLEKPCRENKHTL
HKEIKLKs"
complement(462.. .1661)
/gene="tufA"
/note="synonym: jhp1128"
/complement(462.. .1661)
/gene="tufA"
/note="similar to H. pylori 26695 gene HP1205"
/codon_start=1
/transl_table=11
/product="ELONGATION FACTOR TU (EF-TU)"
/protein_id="AAD06711.1"
/db_xref="GI:4155731"
/translation="MAKEFNRTKPHVNIPTIGHVGHGKTTLSAISAIVLSLUGLAEM
KDYNDIDNAPEKRGITTIATSHIYETENRHYAVDCPGHADYVKNMITGAQMDGA
ILVVSADGPMPTREHILLRSQVGVPHIVFLANKQDMDVDDQELLELMEMEVELLSA
YEFDDTPIVAGSALRALEAKAGNKGWKEKVLKMAEVDYSIPTPERDEKTFILM
PVDVFSIAGTGTVTGIRGVGVKGVDEIVGIRQTKTVTGTVGVNFKELKGEA
GDNVGLLRGTTKKEVERGMVLCKPGSITPHKKPEGEIYVLSKEGGRTPTFTNRP
QFYVRTDVTGTSITLPEGVEMVMPGDNVKITVELISVALELGTKFAIREGGRTVAG
VVSNIIE"

```

gene	2365. .4101 /gene="jhp1129" 2365. .4101 /note="jhp1129" /note="similar to H. pylori 26695 gene HP1206" /codon_start=1 /transl_table=11 /product="abc transporter, ATP-binding protein" /protein_id="AAD06712.1" /db_xref="GI:4155732"	gene	7507. .7602 /gene="jhp1134" 7507. .7602 /note="similar to H. pylori 26695 gene HP1211" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06716.1" /db_xref="GI:4155736"
CDS	<pre>FEVMSLTLLMPEITLASDPSALDDKQWVDFHFSSPVRLAVFSEFCVLGVLFP RMFYGVFTYLRGRENKKAYQIKQQLFLOHLKNYLSHLNHLSDLRDIINNAEGM EMAFNAPLSLLTEIVVFVFTSLITDNKNHEAHLFGNSRKAHDTEIYSTLQWVP EWAASKAQLKSKFNSFNKFTIKDNHEAHLFGNSRKAHDTEIYSTLQWVP YSITVGFSLKAVILFKYGEARVLPITISMYALYRILPSVGTGVSYYNEIA YNQLATVVFSLKSTIVIEEDVPLDFNEKIFLONISFAYKSKHPVLNFNFTIQKGQ KIALIGHSCGKSTLADIIMGLTPKSGEIPIDNTLLSENRRSRKKIGYIPQNYL FDGTVGDNIAGSADDEKRLIKVCKMAHYDFLCEHGLKTQVGGGAKLSGGQKQRI GIARALYDNPEILVDEATSAUDNETESKIMDEIYQIAKNKTLIVIAHRLSTIERCEV IIDMSQHKNLG"</pre>	gene	7507. .7602 /gene="jhp1134" 7507. .7602 /note="similar to H. pylori 26695 gene HP1211" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06716.1" /db_xref="GI:4155736"
gene	complement(4110. .4778) /gene="jhp1130" complement(4110. .4778) /gene="jhp1130" /note="similar to H. pylori 26695 gene HP1207" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06713.1" /db_xref="GI:4155733"	gene	complement(7804. .8121) /gene="atpg" /note="synonym: jhp1135" complement(7804. .8121) /gene="atpg" /note="similar to H. pylori 26695 gene HP1212" /codon_start=1 /transl_table=11 /product="ATP synthase F0, subunit c" /protein_id="AAD06717.1" /db_xref="GI:4155737"
CDS	<pre>VFHYQSGISRNKEIQFYNEILKTPIAQEEIDALAEFGAIEQKLFDRGHLNBEVM AFIDHYQNYIFHIAAALHSELQVLCFEGITKYFKSVESGPPDPKPIIANIIQKYA YDPSRLMIGDSVNDYESAKANKVAFGLYNSKVLNVLQGDQYQCKYLESFKGFDLQ FAKE"</pre>	CDS	<pre>FGALIGMGNAAANTITIGARNPGVGKLTTFVAVAMEAEQVITLVPAILIAYSNP FLS"</pre>
gene	complement(4966. .5955) /gene="M.HpyI" /note="synonym: jhp1131" complement(4966. .5955) /gene="M.HpyI" /note="similar to H. pylori 26695 gene HP1208" /codon_start=1 /transl_table=11 /product="TYPE II DNA MODIFICATION ENZYME (METHYLTRANSFERASE)" /protein_id="AAD06714.1" /db_xref="GI:4155734"	gene	complement(8254. .10320) /gene="pnp" /note="synonym: jhp1136" complement(8254. .10320) /gene="pnp" /note="similar to H. pylori 26695 gene HP1213" /codon_start=1 /transl_table=11 /product="POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE" /protein_id="AAD06718.1" /db_xref="GI:4155738"
CDS	<pre>FKKAVKVISNDLEYSVFLNQNYIGNIQEIIPNKEELINKINSVALKKGFIYSHYSLG GSSRQYFSETNAQKIDAMRLKTEELKSONIDNHSYFFLLASLLESADKVANTASVYG APLKKLKSQAKELILKGAHFDVNLNANEVQODSNDLIGKISGDLILDPYNARQY GANYHLANTIAAYTPTPKGTDLPYSQKSSPFCRSFOILNAPENLIKARPKYIFLSY NNEGLMSETEIKNLIKKGAYSLVTYTNRFKADNKAHKAHVTKELCHLVLIK"</pre>	CDS	<pre>REPSVEDEPLVWVQLEKSYAAGKIPGGVKEGQAQFEILTLRLTLPLPDKD YRYPTQITLMVNHDIENDLOVSALNAASATFLAHAPIKSVSACRIARVNEFTIN PSAIIINOSLDLFSVGTGKESLMIEMESLQKNALEPLMLEALEAKSKIKETCA LYEAAFTPYQNELLPKEGEGIVLNERLLDLKKNQYFDEIIKIGIESALSALSERNVPEK AKKISEAHSEFSLEIEISLEKVKKTEIRMIQDKIRPDKKALBEVRPISIESNLLP</pre>
gene	complement(6284. .6463) /gene="jhp1132" complement(6284. .6463) /gene="jhp1132" /codon_start=1 /transl_table=11 /product="putative" /protein_id="AAD06705.1" /db_xref="GI:4155725"	gene	complement(7804. .8121) /gene="atpg" /note="synonym: jhp1135" complement(7804. .8121) /gene="atpg" /note="similar to H. pylori 26695 gene HP1210" /codon_start=1 /transl_table=11
CDS	<pre>GKVEYKNGIHKRTY"</pre>	CDS	<pre>complement(6692. .7207) /gene="cy88" /note="synonym: jhp1133" complement(6692. .7207) /gene="cy88" /note="similar to H. pylori 26695 gene HP1210" /codon_start=1 /transl_table=11</pre>

Query Match 100.0%; Score 31; DB 1; Length 15292;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAATATTATTCAATAATATTGCTCAAC 31
|||
Db 455 GCTAATATTATTCAATAATATTGCTCAAC 485
|||

RESULT 7
BD061867/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

BD061867 968 bp DNA linear PAT 27-AUG-2002
Antigenic composition and method of detection for Helicobacter pylori.
BD061867.1 GI:23607472
JP 2001517091-A/201.
synthetic construct
synthetic construct
other sequences: artificial sequences.
1 (bases 1 to 968)
Chow, T.P., Fry, K.E., Lim, M.Y. and Mcatee, C.P.
Antigenic composition and method of detection for Helicobacter

```

JOURNAL      Patent: JP 2001517091-A 201 02-OCT-2001;
GENELABS     TECHNOLOGIES INC
COMMENT      PN  JP 2001517091-A/201
PD  02-OCT-2001
PF  25-APR-1998  JP 1998547263
PR  25-APR-1997  US  60/045107,14-OCT-1997 US  60/061958 PI
THERESA P CHOW,KIRK E FRY,MOON Y LIM,C P MCATEE PC
C12N15/31,C07K14/205,C07K16/12,A61K39/106
CC  Strandedness: Single;
CC  Topology: Linear;
FH  Key      Location/Qualifiers.
FEATURES     Location/Qualifiers
source       1..968
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"

ORIGIN
Query Match      94.8%; Score 29.4; DB 6; Length 968;
Best Local Similarity 96.8%; Pred. No. 5.1;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1  GCTAATATTATTCAATAATATTGCTCAAC 31
    |||||
Db  722 GCTAATATTATTCAATAATATTGCTCAAC 692
    |||||

RESULT 8
BD061998/c
LOCUS      BD061998      970 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION Antigenic composition and method of detection for Helicobacter
            Pylori
ACCESSION  BD061998
VERSION    BD061998.1 GI:22607603
KEYWORDS   JP 2001517091-A/332.
SOURCE     synthetic construct
ORGANISM   other sequences; artificial sequences.
REFERENCE  1 (bases 1 to 970)
AUTHORS    Chow,T.P., Fry,K.E., Lim,M.Y. and Mcatee,C.P.
TITLE      Antigenic composition and method of detection for Helicobacter
JOURNAL    Patent: JP 2001517091-A 332 02-OCT-2001;
GENELABS   TECHNOLOGIES INC
COMMENT    PN  JP 2001517091-A/332
PD  02-OCT-2001
PF  25-APR-1998  JP 1998547263
PR  25-APR-1997  US  60/045107,14-OCT-1997 US  60/061958 PI
THERESA P CHOW,KIRK E FRY,MOON Y LIM,C P MCATEE PC
C12N15/31,C07K14/205,C07K16/12,A61K39/106
CC  Strandedness: Single;
CC  Topology: Linear;
FH  Key      Location/Qualifiers.
FEATURES     Location/Qualifiers
source       1..970
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"

ORIGIN
Query Match      94.8%; Score 29.4; DB 6; Length 970;
Best Local Similarity 96.8%; Pred. No. 5.1;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1  GCTAATATTATTCAATAATATTGCTCAAC 31
    |||||
Db  724 GCTAATATTATTCAATAATATTGCTCAAC 694
    |||||

RESULT 9
AX793134
LOCUS      AX793134      30 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Sequence 5598 from Patent WO02066501.
ACCESSION  AX793134

```

```

AX793134.1  GI:32958581
KEYWORDS   Helicobacter pylori
SOURCE     Helicobacter pylori
ORGANISM   Helicobacter pylori
REFERENCE  1
AUTHORS    Legrain,P., Rain,J.C., Colland,F., de Reuse,H. and Labigne,A.
TITLE      Protein-protein interactions in Helicobacter pylori
JOURNAL    Patent: WO 0206501-A 5598 29-AUG-2002;
            Hybrigenics (FR) ; INSTITUT PASTEUR (FR)
FEATURES     Location/Qualifiers
source       1..30
            /organism="Helicobacter pylori"
            /mol_type="unassigned DNA"
            /db_xref="taxon:210"

ORIGIN
Query Match      81.9%; Score 25.4; DB 6; Length 30;
Best Local Similarity 96.3%; Pred. No. 2.8e+02;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  CTAATATTATTCAATAATATTGCTCAC 28
    |||||
Db  4  CTACTATTATTCAATAATATTGCTCAC 30
    |||||

RESULT 10
AC102439
LOCUS      AC102439      184688 bp      DNA      linear      HTG 24-MAR-2004
DEFINITION Mus musculus chromosome 14 clone RP24-14314 map 14, *** SEQUENCING
            IN PROGRESS ***, 3 unordered pieces.
ACCESSION  AC102439
VERSION    AC102439.3 GI:45680590
KEYWORDS   HTG; HTGS PHASE1; HTGS FULLTOP; HTGS ACTIVEFIN.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 184688)
AUTHORS    Birren,B., Nussbaum,C. and Lander,E.
TITLE      Mus musculus chromosome 14, clone RP24-14314
JOURNAL    Unpublished
COMMENT    2 (bases 1 to 184688)
            Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Canarata,J., Campiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,B., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Katat,A., Karatas,A., Kells,C., Larocque,K.,
            Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
TITLE      Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
JOURNAL    Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE  3 (bases 1 to 184688)
AUTHORS    Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,

```

Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K.G., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Riese, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-MAR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 24, 2004 this sequence version replaced gi:22380966.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

Center project name: L18796
Center clone name: 143_I_4

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will be preserved.

* 1 24577: contig of 24577 bp in length
* 24578 24677: gap of 100 bp
* 24678 97652: contig of 72975 bp in length
* 97653 97752: gap of 100 bp
* 97753 184688: contig of 86936 bp in length.

Location/Qualifiers
1. .184688
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/map="14"
/clone="RP24-143I4"
/clone_lib="RPC1-24 Male Mouse BAC"

ORIGIN
Query Match 74.8%; Score 23.2; DB 2; Length 184688;
Best Local Similarity 89.3%; Pred. No. 1.6e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AATATTATTCATATATTTGCTCAAC 31
|||||
Db 62965 AGTACTATTCATATAAATGCTCAAC 62992
|||||

RESULT 11
AC107741/c 212995 bp DNA linear ROD 28-JAN-2004
LOCUS
AC107741 Mus musculus chromosome 3, clone RP23-335P23, complete sequence.
ACCESSION
AC107741
VERSION
AC107741.15 GI:41351636
KEYWORDS
HTC.

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 212995)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus chromosome 3, clone RP23-335P23
Unpublished
2 (bases 1 to 212995)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Kiley, R., Riese, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 212995)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Retta, R., Riese, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (26-NOV-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 212995)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,

Rachupka A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Ronan, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,
 Vasiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (28-JAN-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jan 28, 2004 this sequence version replaced gi:38524693.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20348

Center clone name: 335_P_23

FEATURES

source

Location/Qualifiers

1..212995
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="3"
 /map="3"
 /clone="RP23-335P23"
 /clone_lib="RPCI-23 Female Mouse BAC"

misc_feature

1..20010
 /note="wgs end extension
 clone_end:SP6"

unsure

34..39

/note="<30 qual SNGL region"

complement(259..381)

repeat_region

/rpt_family="MIR"

repeat_region

complement(430..587)

repeat_region

/rpt_family="ORR1D"

repeat_region

complement(930..1069)

repeat_region

1527..1658

repeat_region

/rpt_family="CA)n"

repeat_region

1670..1729

repeat_region

/rpt_family="CA)n"

repeat_region

1774..1918

repeat_region

/rpt_family="CA)n"

repeat_region

1998..2043

repeat_region

/rpt_family="CA)n"

repeat_region

2377..2476

repeat_region

/rpt_family="MIR3"

repeat_region

3457..3523

repeat_region

/rpt_family="t-rich"

repeat_region

3715..3738

repeat_region

/rpt_family="T)n"

repeat_region

complement(4574..4719)

repeat_region

/rpt_family="B1_MM"

repeat_region

complement(4796..4886)

repeat_region

/rpt_family="PB1D10"

repeat_region

complement(4823..4908)

repeat_region

/rpt_family="ID_B1"

repeat_region

complement(4912..4981)

repeat_region

/rpt_family="ID3"

repeat_region

5832..6107

repeat_region

/rpt_family="L1MC2"

repeat_region

6108..6230

repeat_region

/rpt_family="CA)n"

repeat_region

6231..6325

repeat_region

/rpt_family="L1MC2"

repeat_region

6398..6420

repeat_region

/rpt_family="TTTG)n"

repeat_region

complement(6972..7149)

repeat_region
 /rpt_family="Lx4"
 7278..7558
 /rpt_family="B4"
 7630..7655
 /rpt_family="(A)n"
 8955..8992
 /rpt_family="(TTA)n"
 complement(8996..9212)
 /rpt_family="ID_B1"
 12006..12062
 /note="single clone coverage"
 complement(13110..13443)
 /rpt_family="MTC"
 13602..13852
 /rpt_family="ORR1D"
 13853..14253
 /rpt_family="MTA"
 14254..14329
 /rpt_family="ORR1D"
 14368..14558
 /rpt_family="MIR"
 16132..16240
 /rpt_family="MIR"
 16740..16893
 /rpt_family="(TG)n"
 17402..17453
 /rpt_family="(TGAG)n"
 17760..17950
 /rpt_family="ID_B1"
 17964..18160
 /rpt_family="B3"
 complement(18498..18792)
 /rpt_family="B4"
 19418..19438
 /rpt_family="AT-rich"
 19504..19796
 /rpt_family="B4"
 20011..20016
 /note="clone boundary
 clone_end:SP6"
 site:ECORI"
 20450..20737
 /rpt_family="B4"
 20741..20796
 /rpt_family="(CA)n"
 21001..21159
 /rpt_family="L1MB1"
 21252..21304
 /rpt_family="CT-rich"

Query Match 74.8%; Score 23.2; DB 10; Length 212995;

Best Local Similarity 89.3%; Pred. No. 1.5e+02;

Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATGCTCACA 30

Db 22532 TAATATTATTCAATAATATGTTTAAAA 22505

RESULT 12

AC017240

LOCUS

AC017240

DEFINITION

Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.

AC017240

VERSION

AC017240.1 GI:6553746

KEYWORDS

HTG; HTGS PHASE2.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 120794)

AUTHORS

Adams, M. and Venter, J.C.

AC017240 AC017240 120794 bp DNA linear HTG 09-DEC-1999

TITLE
JOURNAL
COMMENT
This sequence was identified as CDM:1020984 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source

1. 120794
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN

Query Match 73.5%; Score 22.8; DB 2; Length 120794;
Best Local Similarity 92.3%; Pred. No. 2.4e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CTAATATTATTCAATAATATTGCTCA 27
||||| ||||||| ||||||| ||||||| |||||||
Db 53851 CTAATAGTATTCAAAAATATTGCTCA 53876

RESULT 13
AC009741/c
LOCUS

DEFINITION
Drosophila melanogaster, chromosome 3R, region 87B-87B, BAC clone
BAC44K17, complete sequence.

ACCESSION

AC009741

VERSION

AC009741.5 GI:13096041

KEYWORDS

HTG.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 166863)
Celniker, S.E., Adams, C.A., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, J.F., An, H., Baldwin, D., Bonzon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Shanahan, M., Strong, R., Svirska, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 3R, region 87B-87B
Unpublished

TITLE
JOURNAL
REFERENCE

2 (bases 1 to 166863)
Celniker, S.E., Agbavani, A., Arcaïna, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E.,
Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

TITLE
JOURNAL

Submitted (30-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT

On Feb 22, 2001 this sequence version replaced gi:5912616.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

FEATURES
source

1. 166863
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strains="Y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="87B-87B"
/clone="BACR44K17 (D976)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACE3.6)"

ORIGIN

Query Match 73.5%; Score 22.8; DB 3; Length 166863;
Best Local Similarity 92.3%; Pred. No. 2.2e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CTAATATTATTCAATAATATTGCTCA 27
||||| ||||||| ||||||| ||||||| |||||||
Db 108454 CTAATAGTATTCAAAAATATTGCTCA 108429

RESULT 14

AE003695/c

LOCUS

DEFINITION
Drosophila melanogaster chromosome 3R, section 33 of 118 of the
complete sequence.

ACCESSION AE003695 AE002708 AE014297

VERSION

AE003695.2 GI:23171087

KEYWORDS

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 225655)
Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L.,
Abail, J.F., Agbavani, A., An, H.J., Andrews-Pfannkuch, C., Baldwin, D.,
Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,
Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S.,
Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P.,
Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A.,
Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,
Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I.,
Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S.,
Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C.,
Ferreira, S., Fleischmann, W., Fosler, C., Gabriellian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennisson, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mates, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nuskern, D.R., Pacleb, J.M., Palazolo, M., Pittman, G.S., Pan, S.,
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R.,

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.	gene
	The genome sequence of <i>Drosophila melanogaster</i>	
	Science 287 (5461), 2185-2195 (2000)	
	20196006	
	10731132	
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	2 (bases 1 to 225655)	CDS
	Celniker, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., Hodgson, A., George, R.A., Hoskins, R.A., Laverty, T., Muzny, D.M., Nelson, C.R., Pacleb, J.M., Park, S., Pfeiffer, B.D., Richards, S., Sodergren, E.J., Svirskas, R., Tabor, P.E., Wan, K., Stapleton, M., Sutton, G.G., Venter, C., Weinstock, G., Scherer, S.E., Myers, E.W., Gibbs, R.A. and Rubin, G.M.	
	Finishing a whole-genome shotgun: release 3 of the <i>Drosophila melanogaster</i> euchromatic genome sequence	
	Genome Biol. 3 (12), RESEARCH0079 (2002)	
	22426065	
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	3 (bases 1 to 225655)	gene
	Misra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradecky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochnik, S.E., Smith, C.D., Tupy, J.L., Whitfield, E.J., Bayraktaroglu, L., Berman, B.P., Bettencourt, B.R., Celniker, S.E., de Grey, A.D., Drysdale, R.A., Harris, N.L., Richter, J., Russo, S., Schroeder, A.J., Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M., Rubin, G.M. and Lewis, S.E.	
	Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a systematic review	
	Genome Biol. 3 (12), RESEARCH0083 (2002)	
	22426069	
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	4 (bases 1 to 225655)	CDS
	Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M. and Celniker, S.E.	
	The transposable elements of the <i>Drosophila melanogaster</i> euchromatin: a genomics perspective	
	Genome Biol. 3 (12), RESEARCH0084 (2002)	
	22426070	
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	5 (bases 1 to 225655)	gene
	Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.	
	Direct Submission	
	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA	
	6 (bases 1 to 225655)	
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	FlyBase	CDS
	Direct Submission	
	Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA	
	7 (bases 1 to 225655)	
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	FlyBase	CDS
	Direct Submission	
	Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA	
	On Sep 18, 2002 this sequence version replaced gi:7299572.	
	Location/Qualifiers	
COMMENT FEATURES source	1..225655	gene
	/organism="Drosophila melanogaster"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:7227"	
	/chromosome="3R"	
tRNA	2735..2806	CDS
	/locus_tag="CR31356"	
	/notes="CR31356-RA"	
	/map="87B2-87B2"	

/protein_id="AAF54761.1"
/db_xref="GI:7299576"
/db_xref="FLYBASE:FBgn0037999"
/translation="MOQLIRVARTLQGRSSGAWGARIAGGNRGIACLAALSETHOIL
OKTREFANAEAPKARHDEELYPAEOVRRLGELGLMSVTVREYVGGSLDYQAYA
IGMEVARGDAVSIUMGVNNLYLCAVQHQTEQKQDFLVPYTCGEHIAFYALSEPG
NGSDAGAASTTAKLQGDSDYINGTKAMISNSKEASGGVFAIVDKSMKHGKITAPLTP
KDFGLSTAKESKMGMRATSTCOLVEDVHVRPSRVLGAAAGDGFKIAMQSLDCGRIG
IAAQATGIAQAALALAVAFGKHARLQIQOKLADMATRVEISRLLTWRAA
WLKNGLPITKEAAMAKLHASESATFCAHQCIQILGGNGYTTDLPAELYRNARVTEI
YECTSEIQRIIVIANAVRELCKE"

Query Match 73.5%; Score 22.8; DB 3; Length 225655;
Best Local Similarity 92.3%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCAATAATATTGCTCA 27
Db 139749 CTAATAGTATTCAAAAATATTGCTCA 139724

RESULT 15
BX571856_06/C
WPCOMMENT
Sequence split into 29 fragments LOCUS BX571856 Accession BX571856

Fragment Name	Begin	End
BX571856_00	1	110000
BX571856_01	100001	210000
BX571856_02	200001	310000
BX571856_03	300001	410000
BX571856_04	400001	510000
BX571856_05	500001	610000
BX571856_06	600001	710000
BX571856_07	700001	810000
BX571856_08	800001	910000
BX571856_09	900001	1010000
BX571856_10	1000001	1110000
BX571856_11	1100001	1210000
BX571856_12	1200001	1310000
BX571856_13	1300001	1410000
BX571856_14	1400001	1510000
BX571856_15	1500001	1610000
BX571856_16	1600001	1710000
BX571856_17	1700001	1810000
BX571856_18	1800001	1910000
BX571856_19	1900001	2010000
BX571856_20	2000001	2110000
BX571856_21	2100001	2210000
BX571856_22	2200001	2310000
BX571856_23	2300001	2410000
BX571856_24	2400001	2510000
BX571856_25	2500001	2610000
BX571856_26	2600001	2710000
BX571856_27	2700001	2810000
BX571856_28	2800001	2902619

Continuation (7 of 29) of BX571856 from base 600001 (BX571856 Staphylococcus aureus subsp

Query Match 72.9%; Score 22.6; DB 1; Length 110000;
Best Local Similarity 86.2%; Pred. No. 2.9e+02;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCAATAATATTGCTCACAA 30
Db 16155 CTTATTTATTATAACATTGCTCACAA 16127

Search completed: April 24, 2005, 03:21:56
Job time : 806.69 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 10:03:10 ; Search time 203.683 Seconds
(without alignments)
900.968 Million cell updates/sec

Title: US-10-039-183A-17

Perfect score: 31

Sequence: 1 gctaattattcaataaattgtcacaac 31

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980a:*
- 2: Geneseqn1990a:*
- 3: Geneseqn2000a:*
- 4: Geneseqn2001a:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002a:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003a:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004a:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	2 AAV07972	AAV07972 Helicobac
2	31	100.0	31	2 AAV07970	AAV07970 Helicobac
C 3	31	100.0	1448	2 AAV07964	AAV07964 Helicobac
C 4	29.4	94.8	968	2 AAV90742	AAV90742 Nucleotid
C 5	29.4	94.8	970	2 AAV90873	AAV90873 Nucleotid
6	25.4	81.9	30	6 ABX69371	ABX69371 Novel Hel
C 7	24	77.4	1200	4 AAS53681	AAS53681 Helicobac
C 8	24	77.4	1200	8 ACA34860	ACA34860 Prokaryot
9	21.6	69.7	30	2 AAV07968	AAV07968 Helicobac
10	21	67.7	2595	6 ABN79830	ABN79830 Fungal ZB
11	21	67.7	2595	6 ABN79831	ABN79831 Fungal ZB
C 12	21	67.7	5333	8 ABZ10049	ABZ10049 Haematopo
C 13	21	67.7	5333	8 ABZ10195	ABZ10195 Haematopo
C 14	21	67.7	8333	4 AAS45406	AAS45406 Chemical
C 15	21	67.7	8333	6 ABL33502	ABL33502 Human imm
C 16	21	67.7	8333	6 ABK28255	ABK28255 DNA trans
17	20.6	66.5	110000	10 ADF77343	ADF77343 Lactic ac
18	20.2	65.2	2169	12 ADQ08665	ADQ08665 Ciona int
19	20.2	65.2	3536	3 AAA79704	AAA79704 Eucalyptu
20	20	64.5	2640	4 AAH76178	AAH76178 A. thalia

21	20	64.5	2671	4 AAH76176	AAH76176 A. thalia
C 22	20	64.5	3129	13 ADR08103	ADR08103 Full leng
C 23	20	64.5	5892	4 ABL12454	ABL12454 Drosophil
C 24	20	64.5	8700	6 ABL33002	ABL33002 Human imm
C 25	20	64.5	11493	5 AAH20175	AAH20175 Mouse epa
C 26	20	64.5	74788	11 ACN45118	ACN45118 Human gen
C 27	20	64.5	110000	6 ABA92787_3	Continuation (4 of
C 28	20	64.5	110000	6 ABA90521_11	Continuation (12 o
C 29	19.8	63.9	228	6 ABN70726	ABN70726 Streptoco
C 30	19.8	63.9	240	6 ABN71501	ABN71501 Streptoco
31	19.8	63.9	507	2 AAV99381	AAV99381 cDNA enco
32	19.8	63.9	1827	10 ADJ56509	ADJ56509 Human cDN
33	19.8	63.9	2153	2 AAQ06168	AAQ06168 Encodes H
34	19.8	63.9	2153	2 AAQ32854	AAQ32854 BMPs, 3/2
35	19.8	63.9	2153	2 AAQ23679	AAQ23679 Encodes H
36	19.8	63.9	2153	2 AAQ41295	AAQ41295 Human BMP
37*	19.8	63.9	2153	2 AAQ99262	AAQ99262 DNA encod
38	19.8	63.9	2153	6 ABK64540	ABK64540 Human ben
39	19.8	63.9	2153	10 ACF05922	ACF05922 Human bon
40	19.8	63.9	2153	10 ACA56729	ACA56729 Human sig
41	19.8	63.9	2153	12 ADH11561	ADH11561 Human bon
42	19.8	63.9	2153	12 ADI56525	ADI56525 Human pol
43	19.8	63.9	4019	9 ACH04010	ACH04010 Human cDN
44	19.8	63.9	110000	6 ABN71527_11	Continuation (12 o
45	19.8	63.9	225734	12 ADQ59377	ADQ59377 Human can

ALIGNMENTS

RESULT 1

AAV07972

ID AAV07972 standard; DNA; 31 BP.

XX AAV07972;

AC AAV07972;

XX 25-MAR-2003 (revised)

DT 02-FEB-1999 (first entry)

XX Helicobacter pylori polypeptide GHPO 750 3' DNA primer.

DE GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;

KW primer; ss.

XX Synthetic.

OS Helicobacter pylori.

XX WO9843479-Al.

PN 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

PF 01-APR-1997; 97US-00831310.

XX 01-APR-1997; 97US-00834666.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

PI WPI; 1998-568251/48.

DR New isolated Helicobacter polynucleotides - used to develop products for

XX gastroduodenal infections and

PT Claim 5; Page 156; 184pp; English.

XX This 3' primer is used with a 5' primer (see AAV07971) in the PCR

CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic DNA in

XX order to obtain DNA (see AAV07964) encoding the a 50 kDa polypeptide (see

CC AAW73035) designated GHPO 750. The isolated polynucleotide, and encoded

CC polypeptide, can be used to develop vaccines for the treatment and
CC prevention of Helicobacter infections. (Updated on 25-MAR-2003 to correct
CC PI field.)
XX

SQ Sequence 31 BP; 12 A; 6 C; 2 G; 11 T; 0 U; 0 Other;

Query Match 100.0%; Score 31; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAATATTATTCATAATATTGCTCACAAC 31
|||||
Db 1 GCTAATATTATTCATAATATTGCTCACAAC 31

RESULT 2

AAV07970
ID AAV07970 standard; DNA; 31 BP.

XX AAV07970;

XX 25-MAR-2003 (revised)
DT 02-FEB-1999 (first entry)

XX Helicobacter pylori polypeptide GHPO 1360 5' DNA primer.

DE GHPO 1360; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;
KW primer; ss.

XX Synthetic.
OS Helicobacter pylori.

XX WO9843479-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

XX 01-APR-1997; 97US-00831310.
PR 01-APR-1997; 97US-00834666.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PA (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.

XX New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastroduodenal diseases.

PS Claim 5; Page 156; 184pp; English.

CC This 5' primer is used with a 3' primer (see AAV07970) in the PCR
CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic DNA in
CC order to obtain DNA (see AAV07963) encoding the mature form of a 32 kDa
CC polypeptide (see AAW73034) designated GHPO 1360. The isolated
CC polynucleotide, and encoded polypeptide, can be used to develop vaccines
CC for the treatment and prevention of Helicobacter infections. (Updated on
CC 25-MAR-2003 to correct PI field.)
XX

SQ Sequence 31 BP; 12 A; 6 C; 2 G; 11 T; 0 U; 0 Other;

Query Match 100.0%; Score 31; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAATATTATTCATAATATTGCTCACAAC 31
|||||
Db 1 GCTAATATTATTCATAATATTGCTCACAAC 31

RESULT 3

AAV07964/c
ID AAV07964 standard; DNA; 1448 BP.

XX AAV07964;

XX 25-MAR-2003 (revised)
DT 02-FEB-1999 (first entry)

XX Helicobacter pylori 50 kDa polypeptide GHPO 750 DNA.

DE GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; ss.
KW Helicobacter pylori.

XX Key Location/Qualifiers
FH 118..1317
FT CDS /*tag= a

XX WO9843479-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

XX 01-APR-1997; 97US-00831310.

XX 01-APR-1997; 97US-00834666.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PA (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.
DR P-PSDB; AAW73035.

XX New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastroduodenal diseases.

PS Claim 1; Page 150-152; 184pp; English.

XX This DNA sequence codes for a 50 kDa Helicobacter pylori polypeptide (see
CC AAW73035) designated GHPO 750. A polynucleotide encoding GHPO 750 can be
CC obtained from genomic DNA by PCR amplification (see AAV07971-72). The
CC invention provides polynucleotides (see AAV72001, AAV07912-21 and
CC AAV70963-64) encoding a family 76 kDa Helicobacter polypeptides (see
CC AAW73022-32), GHPO 750 and a 32 kDa polypeptide (see AAW73034). These
CC polynucleotides were initially identified in a search of H. pylori
CC genomic databases. DNA cassettes for expression of the Helicobacter
CC proteins (unprocessed or mature forms) in prokaryotic or eukaryotic cells
CC are provided. The polynucleotides can be used in vaccines to prevent or
CC treat Hb infection in a mammal. Viral (especially poxvirus) or bacterial
CC vectors are used. Products and methods of the invention allow treatment
CC and prevention of gastroduodenal diseases associated with Hb infections,
CC including acute, chronic, and atrophic gastritis, and peptic ulcer
CC diseases, e.g. gastric and duodenal ulcers. Diagnostic and detection
CC methods are also provided. GHPO 750 was demonstrated to be a protective
CC antigen. (Updated on 25-MAR-2003 to correct PI field.)
XX

SQ Sequence 1448 BP; 466 A; 237 C; 364 G; 381 T; 0 U; 0 Other;

Query Match 100.0%; Score 31; DB 2; Length 1448;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAATATTATTCATAATATTGCTCACAAC 31
|||||

Db 1324 GCTAATATTATTCATAATATTGCTCACAAC 1294

RESULT 4

AAV90742/c

```

ID AAV90742 standard; DNA; 968 BP.
XX AC
XX AAV90742;
XX DT
XX 20-MAR-2003 (revised)
XX DT 18-FEB-1999 (first entry)
XX DE
XX Nucleotide sequence of clone Z14.ASM from cluster 30.
XX KW
XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX OS
XX Helicobacter pylori.
XX PN
XX WO9849314-A2.
XX PD
XX 05-NOV-1998.
XX PF
XX 25-APR-1998; 98WO-US008487.
XX PR
XX 25-APR-1997; 97US-0045107P.
XX PR 14-OCT-1997; 97US-0061958P.
XX PA
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX PI
XX Chow TP, Fry KE, Lim MY, Mcatee CP;
XX WPI; 1999-009433/01.
XX PS
XX Claim 27; Page 170; 402pp; English.
XX CC
XX The present sequence encodes a Helicobacter pylori antigenic protein that
XX is characterised by immunoreactivity with H. pylori-positive antisera.
XX The proteins are highly immunogenic and induce a long-lasting immune
XX response that persists even after antimicrobial treatment. In antibody-
XX detection assays, on sera, plasma, urine, saliva etc., they are highly
XX sensitive and specific. The specification also describes 69 previously
XX unrecognised immunogenic cluster families. H. pylori antigens are used to
XX detect H. pylori-specific antibodies, for diagnosing infection or to
XX confirm eradication of infection, and in vaccines to protect against H.
XX pylori infection and related diseases (gastritis, peptic ulcer, gastric
XX adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
XX SQ
XX Sequence 968 BP; 306 A; 154 C; 260 G; 248 T; 0 U; 0 Other;
XX
XX Query Match 94.8%; Score 29.4; DB 2; Length 968;
XX Best Local Similarity 96.8%; Pred. No. 0.42;
XX Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
OY 1 GCTATATATTTCATATATATTGCTCACAAAC 31
Db 722 GCTATATATTTCATATATATTGCTCACAAAC 692

RESULT 5
AAV90873/c
ID AAV90873 standard; DNA; 970 BP.
XX AC
XX AAV90873;
XX DT
XX 20-MAR-2003 (revised)
XX DT 18-FEB-1999 (first entry)
XX DE
XX Nucleotide sequence of cluster 30.
XX KW
XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX OS
XX Helicobacter pylori.
XX
XX Query Match 94.8%; Score 29.4; DB 2; Length 968;
XX Best Local Similarity 96.8%; Pred. No. 0.42;
XX Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
OY 1 GCTATATATTTCATATATATTGCTCACAAAC 31
Db 722 GCTATATATTTCATATATATTGCTCACAAAC 692

RESULT 6
ABX69371
ID ABX69371 standard; DNA; 30 BP.
XX AC
XX ABX69371;
XX DT
XX 07-MAY-2003 (first entry)
XX DE
XX Novel Helicobacter pylori gene PCR primer #2342.
XX KW
XX Protein-protein interaction; ulcer; selected interacting domain; SID;
XX KW PCR; primer; ss.
XX OS
XX Helicobacter pylori.
XX PN
XX WO200266501-A2.
XX PD
XX 29-AUG-2002.
XX PF
XX 28-DEC-2001; 2001WO-EP015428.
XX PR
XX 02-JAN-2001; 2001US-0259302P.
XX PA
XX (HYBR-) HYBRIGENICS.
XX PA (INSP) INST PASTEUR.
XX PI
XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX
XX Query Match 94.8%; Score 29.4; DB 2; Length 970;
XX Best Local Similarity 96.8%; Pred. No. 0.42;
XX Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
OY 1 GCTATATATTTCATATATATTGCTCACAAAC 31
Db 724 GCTATATATTTCATATATATTGCTCACAAAC 694

RESULT 6
ABX69371
ID ABX69371 standard; DNA; 30 BP.
XX AC
XX ABX69371;
XX DT
XX 07-MAY-2003 (first entry)
XX DE
XX Novel Helicobacter pylori gene PCR primer #2342.
XX KW
XX Protein-protein interaction; ulcer; selected interacting domain; SID;
XX KW PCR; primer; ss.
XX OS
XX Helicobacter pylori.
XX PN
XX WO200266501-A2.
XX PD
XX 29-AUG-2002.
XX PF
XX 28-DEC-2001; 2001WO-EP015428.
XX PR
XX 02-JAN-2001; 2001US-0259302P.
XX PA
XX (HYBR-) HYBRIGENICS.
XX PA (INSP) INST PASTEUR.
XX PI
XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX

```

DR WPI; 2002-674910/72.
XX New complexes of protein-protein interactions in *Helicobacter pylori*,
PT useful for identifying modulating compounds for treating or preventing
PT ulcers in mammals.
XX
PS Example 9; Page 561; 642pp; English.
XX
CC The invention describes a complex of protein-protein interactions in
CC *Helicobacter pylori* selected from 421 complexes given in the
CC specification. The complex of protein-protein interactions are useful for
CC screening for agents which modulate the interaction of proteins.
CC Modulating compounds which binds to a targeted bacterial protein may be
CC used for treating or preventing ulcers in a human or animal. This
CC sequence represents a primer used to isolate polynucleotides encoding
CC *Helicobacter pylori* proteins for studies on protein-protein interactions
XX
SQ Sequence 30 BP; 10 A; 7 C; 1 G; 9 T; 3 U; 0 Other;
Query Match 81.9%; Score 25.4; DB 6; Length 30;
Best Local Similarity 88.9%; Pred. No. 11;
Matches 24; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 CTATATTATTCAATAATATTGCTCAC 28
|:|:|||||||||||||||||
Db 4 CUACUATTATTCATAATATTGCTCAC 30
RESULT 7
AAS53681/c
ID AAS53681 standard; DNA; 1200 BP.
XX
AC AAS53681;
XX
DT 13-FEB-2002 (first entry)
XX
DE *Helicobacter pylori* DNA for cellular proliferation protein #135.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS *Helicobacter pylori*.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlseen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU35822.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 7318; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes

CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1200 BP; 365 A; 207 C; 319 G; 309 T; 0 U; 0 Other;
Query Match 77.4%; Score 24; DB 4; Length 1200;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 TTATTCATAATATTGCTCACAC 31
|:|:|||||||||||||||||
Db 1200 TTATTCATAATATTGCTCACAC 1177
RESULT 8
ACA34860/c
ID ACA34860 standard; DNA; 1200 BP.
XX
AC ACA34860;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #16517.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS *Helicobacter pylori*.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU30990.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 22730; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation or to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1200 BP; 365 A; 207 C; 319 G; 309 T; 0 U; 0 Other;

Query Match 77.4%; Score 24; DB 8; Length 1200;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 TTATTCATATATATGCTCACAC 31
 |||||
 DB 1200 TTATTCATATATATGCTCACAC 1177

RESULT 9

AAV07968
 ID AAV07968 standard; DNA; 30 BP.
 XX AC AAV07968;
 XX
 XX 25-MAR-2003 (revised)
 DT 02-FEB-1999 (first entry)
 XX
 XX Helicobacter pylori polypeptide GHPO 750 3' DNA primer.
 DE
 XX GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;
 KW primer; ss.
 XX
 XX Synthetic.
 OS Helicobacter pylori.
 XX
 XX WO9843479-A1.
 PN
 XX
 XX 08-OCT-1998.
 PD
 XX 31-MAR-1998; 98WO-US006421.
 PF
 XX 01-APR-1997; 97US-00831310.
 PR 01-APR-1997; 97US-00834666.
 XX
 XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Kleanthous H, Liesolo L, Tomb J, Miller C, Algarawi A;
 PI
 XX WPI; 1998-568251/48.
 DR
 XX New isolated Helicobacter polynucleotides - used to develop products for
 PT the diagnosis, prevention and treatment of Helicobacter infections and

PT gastroduodenal diseases.

XX Example 3.B; Page 64; 184pp; English.

XX This 3' primer was used with a 5' primer (see AAV07967) in the PCR
 CC amplification of Helicobacter pylori strain ORV2001 genomic DNA in order
 CC to obtain DNA (see AAV07964) encoding a 50 kDa polypeptide (see AAV73035)
 CC designated GHPO 750. The primer pair includes a 5' clamp and BamHI and
 CC XhoI restriction enzyme recognition sequences for cloning purposes. The
 CC PCR product was ligated into vector pET28a, and recombinant polypeptide
 CC was expressed as a histidine-tagged fusion protein in *E. coli* host cells.
 CC The polypeptide can be used to develop vaccines for the treatment and
 CC prevention of Helicobacter infections. (Updated on 25-MAR-2003 to correct
 CC PI field.)

XX SQ Sequence 30 BP; 8 A; 8 C; 4 G; 10 T; 0 U; 0 Other;

Query Match 69.7%; Score 21.6; DB 2; Length 30;
 Best Local Similarity 85.7%; Pred. No. 2.4e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GCTAATATTATTCAATATATATGCTCAC 28
 |||||
 DB 3 GCTCGAGTTATTCAATATATATGCTCAC 30

RESULT 10

ABN79830
 ID ABN79830 standard; DNA; 2595 BP.

XX AC ABN79830;

XX 24-JUL-2002 (first entry)

XX Fungal ZBC gene sequence #67.

XX Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;
 KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;
 KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;
 KW angio genesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;
 KW fungal toxin; cell surface receptor; plant growth regulator; pigment;
 KW insecticide; antineoplastic; gene; ds.

XX Unidentified.

XX WO200224865-A2.

XX 28-MAR-2002.

XX 19-SEP-2001; 2001WO-US029288.

XX 19-SEP-2000; 2000US-0233564P.

XX (MICR-) MICROBIA INC.

XX Holtzman D, Madden K, Maxon M, Sherman A;

XX WPI; 2002-352005/38.

XX P-PSDB; ABP35641.

XX New method for improving the production of a secondary metabolite e.g.
 PT antineoplastic agent, ergot alkaloid from a fungus involves modulation of
 PT the expression of at least one zinc binuclear cluster protein gene.

XX Claim 2; SEQ ID NO 145; 49pp + Sequence Listing; English.

XX The invention relates to improving the production of a secondary
 CC metabolite by a fungus. This involves modulating the expression of at
 CC least one ZBC (zinc binuclear cluster protein) gene in a manner to
 CC improve the yield of the secondary metabolite. Methods of the invention
 CC may be used for improving the production of the secondary metabolite e.g.
 CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such
 CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),

an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such as avallin), a glucan synthase inhibitor, gliotoxin family of compounds, a fungal toxin, a modulator of cell surface receptor signalling, a plant growth regulator, a pigment, an insecticide, or an antineoplastic compound. The method results in a decrease in fermentor run-time, a decrease in the size of the fermentor required for the production of equivalent amounts of the secondary metabolite, or a decrease in the biomass required for the production, which translates into decreased waste that must be handled in downstream processing. The sequences given in records ABN79764-ABN79911 represent ZEC genes of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained directly from WIPO at [ftp.wipo.int/pub/published/pct/sequences](http://wipo.int/pub/published/pct/sequences)

XX
SQ Sequence 2595 BP; 829 A; 588 C; 491 G; 687 T; 0 U; 0 Other;

Query Match	67.7%	Score 21;	DB 6;	Length 2595;
Best Local Similarity	82.8%;	Pred. No. 3.8e+02;		
Matches 21;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps	0;			

	increased	217,	conspicuous	67,	nismatches	5,	indels	6,	gaps	6
Qy	2	CTAATATTATTCAATAATATTGCTCACA	30							
Db	1027	CTAATCTAAACCAATAATATTACTCATAA	1055							

RESULT 11	
ABN79831	
ID	ABN79831 standard; DNA; 2595 BP.
XX	
AC	ABN79831;
XX	
DT	24-JUL-2002 (first entry)
XX	
DE	Fungal ZBC gene sequence #68.

XX	Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;
KW	antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastin;
KW	mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;
KW	angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;
KW	fungal toxin; cell surface receptor; plant growth regulator; pigment;
KW	insecticide; antineoplastic; gene; ds.
XX	
XX	Unidentified.
OS	
XX	
PN	WO200224865-A2.
XX	
XX	28-MAR-2002.
PP	
XX	19-SEP-2001; 2001WO-US029288.
XX	
XX	19-SEP-2000; 2000US-0233564P.
PR	
XX	(MICR-) MICROBIA INC.
XX	
PI	Holtzman D, Madden K, Maxon M, Sherman A;
XX	
XX	WPI; 2002-352005/38.
DR	
XX	P-PSDB; ABP35642.
XX	
PT	New method for improving the production of a secondary metabolite e.g.
PT	antineoplastic agent, ergot alkaloid from a fungus involves modulation of
PT	the expression of at least one zinc binuclear cluster protein gene.

SQ Sequence 2595 BP; 829 A; 588 C; 490 G; 688 T; 0 U; 0 Other;

Query Match	67.7%	Score 21;	DB 6;	Length 2595;
Best Local Similarity	82.8%;	Pred. No. 3.8e+02;		
Matches 24;	Conservative	0;	Mismatches 5;	Indels 0;
				Gaps 0;

Qy 2 CTATATTATTCAATAATAATTGCTCAAA 30
||||| ||| ||||| ||| |||
Dp 1027 CTAACTATAACAATAATAATTACTATAA 1055

RESULT 12
ABZ10049/c
ID ABZ10049 standard; DNA; 5333 BP.
XX
XX ABZ10049;
AC AC
XX
XX
16-JAN-2003 (first entry)
DT DT
XX
XX
Haematopoietic cell proliferation disorder related DNA sequence #189.

XX	Human; haematopoietic cell proliferation disorder; cytostatic;
KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW	cytosine methylation state; gene; ds.
XX	
XX	Homo sapiens.
OS	
XX	
PN	WO200277272-A2.
XX	
XX	03-OCT-2002.
XX	
PF	26-MAR-2002; 2002WO-EP003401.
XX	
XX	26-MAR-2001; 2001US-027833P.
PR	
XX	(EPIG-) EPIGENOMICS AG.
PA	
XX	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI	Olek A, Pleepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI	Schwöpe I, Ziebarth H;
XX	
XX	WPI; 2003-018942/01.
DR	
XX	Detecting and differentiating between hematopoietic cell proliferative
PT	disorders, comprises contacting a target nucleic acid with a reagent that
PT	distinguishes between methylated and non-methylated CpG dinucleotides.
XX	
XX	Claim 28; SEQ ID NO 189; 117pp; English.
PS	

The present invention describes a method for detecting and differentiating between haematopoietic cell proliferative disorders associated with at least 1 gene and/or their regulatory regions in a subject. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. AB209061 to AB21118 represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used: for differentiating between healthy haematopoietic cells and proliferative

CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 5333 BP; 1248 A; 87 C; 1254 G; 2744 T; 0 U; 0 Other;
 Query Match 67.7%; Score 21; DB 8; Length 5333;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 CTAATATTATTCATAATATTGCTCACAA 30
 |||||
 DB 2983 CTAATATTATTCATAATATTGCTTAA 2955
 |||||
 RESULT 13
 ID ABZ10195/c
 AC ABZ10195;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #335.
 XX
 KW Human; haematopoietic cell proliferation disorder; cytosine;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20027272-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-EP003401.
 XX
 PR 26-MAR-2001; 2001US-0278333P.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Ley E;
 PI Lewin A, Lipicher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
 PI Schwöbe I, Ziebarth H;
 XX
 DR WPI; 2003-018942/01.
 XX
 PT Detecting and differentiating between haematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.
 XX
 PS Claim 28; SEQ ID NO 335; 117pp; English.
 XX
 CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative

CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 5333 BP; 1248 A; 0 C; 1254 G; 2831 T; 0 U; 0 Other;
 Query Match 67.7%; Score 21; DB 8; Length 5333;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 CTAATATTATTCATAATATTGCTCACAA 30
 |||||
 DB 2983 CTAATATTATTCATAATATTGCTTAA 2955
 |||||
 RESULT 14
 ID AAS45406/c
 AC AAS45406 standard; DNA; 8333 BP.
 XX
 AC AAS45406;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Chemically pretreated genomic DNA associated with cell cycle #56.
 XX
 KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
 KW immunosuppressive; antitumour; cytosine; cytosine; antiarteriosclerotic; ds;
 KW PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200168911-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP002945.
 XX
 PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602751/68.
 XX
 PT Designing primers and probes for analyzing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle.
 XX
 PS Claim 1; SEQ ID NO 111; 28pp; English.
 XX
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated

CC DNA, Genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers

XX SQ Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 4; Length 8333;
Best Local Similarity 82.8%; Pred. No. 3.8e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CTAATATTATTCAATAATATTCCTCACAA 30
|||||
Db 5983 CTAATATTATTAAATAATATTATTCTTAA 5955

RESULT 15

ABL33502/c

ID ABL33502 standard; DNA; 8333 BP.

XX AC ABL33502;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 1475.

XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; anianaemic; cytostatic; nootropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
XX KW antinflamatory; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.

XX OS Homo sapiens.

XX PN WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EF007537.

XX PR 30-JUN-2000; 2000DE-01032529.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130909/17.

XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.

XX PS Claim 1; SEQ ID NO 1475; 32pp + Sequence Listing; German.

XX CC The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention

XX SQ Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 6; Length 8333;
Best Local Similarity 82.8%; Pred. No. 3.8e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CTAATATTATTCAATAATATTCCTCACAA 30
|||||
Db 5983 CTAATATTATTAAATAATATTATTCTTAA 5955

Search completed: April 24, 2005, 02:20:08
Job time : 204.683 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:51:24 ; Search time 60.2535 Seconds
(without alignments)
841.853 Million cell updates/sec

Title: US-10-039-183A-17
Perfect score: 31
Sequence: 1 gctaattattcaataattgctcacaac 31

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21.6	69.7	278866	4	US-09-949-016-13922
C 2	21.6	69.7	278866	4	US-09-949-016-13923
C 3	21.6	69.7	278866	4	US-09-949-016-13924
C 4	21.6	69.7	278866	4	US-09-949-016-13925
C 5	21.6	69.7	278866	4	US-09-949-016-13926
C 6	21.6	69.7	278866	4	US-09-949-016-14699
C 7	21.6	69.7	278866	4	US-09-949-016-14700
C 8	21.6	69.7	278866	4	US-09-949-016-14701
C 9	21.6	69.7	278866	4	US-09-949-016-14702
C 10	21.6	69.7	278866	4	US-09-949-016-14703
C 11	20.4	65.8	94077	4	US-09-949-016-13635
C 12	20.2	65.2	407	4	US-09-422-978-317
C 13	20.2	65.2	3536	4	US-10-101-464A-856
C 14	20	64.5	11492	4	US-09-693-205A-3
C 15	20	64.5	43192	4	US-09-949-016-15466
C 16	20	64.5	58789	4	US-09-949-016-15922
C 17	20	64.5	268449	4	US-09-949-016-15922
C 18	20	64.5	640681	4	US-09-949-016-17244
C 19	19.8	63.9	507	3	US-09-790-988-1
C 20	19.8	63.9	1866	3	US-08-868-452-41
C 21	19.8	63.9	2149	4	US-09-248-796A-5265
C 22	19.8	63.9	2153	1	US-09-949-016-1611
C 23	19.8	63.9	2153	1	US-07-539-756-1
C 24	19.8	63.9	2153	1	US-08-377-292-5
C 25	19.8	63.9	2153	2	US-07-989-847-9
C 26	19.8	63.9	2153	3	US-08-469-411-9
C 27	19.8	63.9	2153	4	US-09-016-434-1327
C 28	19.8	63.9	2153	4	US-09-780-601A-9

28	19.8	63.9	124110	4	US-09-949-016-13353	Sequence 13353, A
29	19.6	63.2	601	4	US-09-949-016-60540	Sequence 60540, A
30	19.6	63.2	601	4	US-09-949-016-60541	Sequence 60541, A
C 31	19.6	63.2	601	4	US-09-949-016-157403	Sequence 157403, A
C 32	19.6	63.2	601	4	US-09-949-016-157510	Sequence 157510, A
C 33	19.6	63.2	59519	4	US-09-949-016-13504	Sequence 13504, A
C 34	19.6	63.2	126176	4	US-09-949-016-16137	Sequence 16137, A
C 35	19.6	63.2	126176	4	US-09-949-016-16138	Sequence 16138, A
C 36	19.4	62.6	260247	4	US-09-949-016-13358	Sequence 13358, A
C 37	19.2	61.9	23094	4	US-09-949-016-13468	Sequence 13468, A
C 38	19	61.3	282	4	US-09-861-451A-51	Sequence 51, Appl
C 39	19	61.3	601	4	US-09-949-016-170105	Sequence 170105, A
C 40	19	61.3	1409	4	US-09-976-594-349	Sequence 349, App
C 41	19	61.3	2054	4	US-09-949-016-11199	Sequence 1199, Ap
C 42	19	61.3	2054	4	US-09-949-016-1200	Sequence 1200, Ap
C 43	19	61.3	2120	4	US-09-949-016-11197	Sequence 1197, Ap
C 44	19	61.3	2120	4	US-09-949-016-11198	Sequence 1198, Ap
C 45	19	61.3	2137	4	US-09-976-594-348	Sequence 348, App

ALIGNMENTS

RESULT 1
US-09-949-016-13922/c
; Sequence 13922, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13922
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13922

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATATTGCTCACA 30
|||||
DB 65820 TAATATTATTCAATATTGCTCAGAA 65793
|||||

RESULT 2
US-09-949-016-13923/c
; Sequence 13923, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13923
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13923

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 3
US-09-949-016-13924/c
; Sequence 13924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13924
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13924

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 4
US-09-949-016-13925/c
; Sequence 13925, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13925
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13925

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 5
US-09-949-016-13926/c
; Sequence 13926, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13926
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13926

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 6
US-09-949-016-14699/c
; Sequence 14699, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14699
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14699

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATGCTCACA 30
Db 65820 TAATATTATTCAATAATATGCTCAGAA 65793

RESULT 7
US-09-949-016-14700/c
; Sequence 14700, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14700
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14700

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATGCTCACA 30
Db 65820 TAATATTATTCAATAATATGCTCAGAA 65793

RESULT 8
US-09-949-016-14701/c
; Sequence 14701, Application US/09949016
```

```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14701
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14701

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATGCTCACA 30
Db 65820 TAATATTATTCAATAATATGCTCAGAA 65793

RESULT 9
US-09-949-016-14702/c
; Sequence 14702, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14702
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14702

Query Match          69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATGCTCACA 30
Db 65820 TAATATTATTCAATAATATGCTCAGAA 65793
```

RESULT 10

US-09-949-016-14703/C
; Sequence 14703, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14703
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14703

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCAAA 30
|||||
DB 65820 TAATATTATTCAATAATATTGCTCAAA 65793

RESULT 11

US-09-949-016-13635
; Sequence 13635, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13635
; LENGTH: 94077
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13635

Query Match 65.8%; Score 20.4; DB 4; Length 94077;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTATATATTCAATAATATTGCTCAAA 30
|||||
DB 45490 GCTATTGTATTCAATATTATTGATTACAA 45519

RESULT 12

US-09-422-978-317/C
; Sequence 317, Application US/09422978
; Patent No. 6537751
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
; FILE REFERENCE: GENSET.020CPI
; CURRENT APPLICATION NUMBER: US/09/422,978
; PRIOR FILING DATE: 1999-10-20
; EARLIER APPLICATION NUMBER: US 09/298,850
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,732
; EARLIER FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: US 60/082,614
; EARLIER FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 317
; LENGTH: 47
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 24
; OTHER INFORMATION: 99-14186-424 : polymorphic base A or G
US-09-422-978-317

Query Match 65.2%; Score 20.2; DB 4; Length 47;
Best Local Similarity 81.5%; Pred. No. 90;
Matches 22; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 29
|||||
DB 45 TAATAGTTTTCTACAATATTGTCACA 19

RESULT 13

US-10-101-464A-856
; Sequence 856, Application US/10101464A
; Patent No. 6768041
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCI/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 856
; LENGTH: 3536
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-101-464A-856

Query Match 65.2%; Score 20.2; DB 4; Length 3536;
Best Local Similarity 88.0%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCA 27
|||||
DB 181 TAATTTTATTCAACAATATTGCTCA 205

RESULT 14

US-09-693-205A-3/c
; Sequence 3, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; CURRENT FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 11492
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-693-205A-3

Query Match 64.5%; Score 20; DB 4; Length 11492;
Best Local Similarity 82.1%; Pred. NO. 1.4e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30
|||||
Db 5105 TAATATTGGCATAAATTTTCTCAAA 5078

RESULT 15

US-09-949-016-15466/c
; Sequence 15466, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15466
; LENGTH: 43192
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15466

Query Match 64.5%; Score 20; DB 4; Length 43192;
Best Local Similarity 82.1%; Pred. NO. 1.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30
|||||
Db 26391 TAAATATTCAATAATATTGGAATAAA 26364

Search completed: April 24, 2005, 05:33:14
Job time : 64.4535 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 02:04:39 ; Search time 250.62 Seconds
(without alignments)
750.949 Million cell updates/sec

Title: US-10-039-183A-17

Perfect score: 31

Sequence: 1 gctaattattcaataattgtctcaac 31

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 5633728 seqs, 303525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	8 US-08-831-310-15	Sequence 15, Appl
2	31	100.0	31	8 US-08-831-310-17	Sequence 17, Appl
3	31	100.0	31	15 US-10-039-183A-15	Sequence 15, Appl
4	31	100.0	31	15 US-10-039-183A-17	Sequence 17, Appl
5	31	100.0	1448	8 US-08-831-310-3	Sequence 3, Appl
6	31	100.0	1448	15 US-10-039-183A-3	Sequence 3, Appl
7	31	100.0	2825	19 US-10-662-126-32	Sequence 32, Appl
8	31	100.0	2825	19 US-10-662-126-32	Sequence 32, Appl
9	24	77.4	1200	9 US-09-815-242-7318	Sequence 7318, Ap
10	24	77.4	1200	17 US-10-282-122A-22730	Sequence 8, Appl
11	21.6	69.7	30	8 US-08-831-310-8	Sequence 8, Appl
12	21.6	69.7	30	15 US-10-039-183A-8	Sequence 8, Appl

Sequence 28120, A
Sequence 145, App
Sequence 147, App
Sequence 189, App
Sequence 335, App
Sequence 113, App
Sequence 1475, App
Sequence 1475, App
Sequence 24648, A
Sequence 6941, Ap
Sequence 25820, A
Sequence 27834, A
Sequence 166130, A
Sequence 166131, A
Sequence 22, Appl
Sequence 57, Appl
Sequence 317, App
Sequence 71415, A
Sequence 856, App
Sequence 856, App
Sequence 3, Appl
Sequence 975, App
Sequence 1906, Ap
Sequence 1, Appl
Sequence 2, Appl
Sequence 315, App
Sequence 692, App
Sequence 435, App
Sequence 5, Appl
Sequence 13, Appl
Sequence 9, Appl
Sequence 1327, Ap
Sequence 2020, Ap

US-10-425-115-28120
US-10-149-310-145
US-10-149-310-147
US-10-473-126-189
US-10-473-126-335
US-10-239-676-113
US-10-311-455-1475
US-10-240-453-129
US-10-425-114-24648
US-10-425-114-6941
US-10-425-114-25820
US-10-425-114-27834
US-10-425-115-166130
US-10-425-115-166131
US-10-739-930-22
US-10-175-523-57
US-10-349-143-317
US-10-424-599-71415
US-10-101-464A-856
US-10-864-252-856
US-10-203-351-3
US-10-311-455-975
US-10-087-192-1906
US-10-810-788A-6
US-09-790-988-1
US-10-312-841-2
US-10-084-817-315
US-09-960-706-692
US-09-873-319-435
US-10-286-152A-5
US-10-366-345-13
US-10-375-150-9
US-10-305-720-1327
US-10-062-674-2020

ALIGNMENTS

RESULT 1

US-08-831-310-15

Sequence 15, Application US/08831310

Publication No. US20020026035A1

GENERAL INFORMATION:

APPLICANT: Kleantous, Harold et al.

TITLE OF INVENTION: Helicobacter GHPO 1360 and

TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Elbing LLP

STREET: 176 Federal Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/831,310

FILING DATE: 01-APR-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,175

REFERENCE/DOCKET NUMBER: 06132/037001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200

TELEFAX: 617-428-7045

INFORMATION FOR SEQ ID NO: 15:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-831-310-15
Query Match 100.0%; Score 31; DB 8; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
Db 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
RESULT 2
US-08-831-310-17
; Sequence 17, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-831-310-17
Query Match 100.0%; Score 31; DB 8; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
Db 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
RESULT 3
US-10-039-183A-15
; Sequence 15, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
```

```
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-15
Query Match 100.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
Db 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
RESULT 4
US-10-039-183A-17
; Sequence 17, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-17
Query Match 100.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
Db 1 GCTAATATTATTCAATAATATTGCTCACAAAC 31
RESULT 5
US-08-831-310-3/c
; Sequence 3, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
```

Query Match 100.0%; Score 31; DB 15; Length 1448;
Best Local Similarity 100.0%; Pred. No. 0.33;

;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 7318
;; LENGTH: 1200
;; TYPE: DNA
;; ORGANISM: Helicobacter pylori
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1200)
US-09-815-242-7318

Query Match 77.4%; Score 24; DB 9; Length 1200;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TTATTCAATAATATTGCTCACAAAC 31
Db 1200 TTATTCAATAATATTGCTCACAAAC 1177

RESULT 9

US-10-282-122A-22730/c
;; Sequence 22730, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Lianguo
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsythe, R.
;; APPLICANT: Xu, H.
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 22730
;; LENGTH: 1200
;; TYPE: DNA
;; ORGANISM: Helicobacter pylori
US-10-282-122A-22730

Query Match 77.4%; Score 24; DB 17; Length 1200;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TTATTCAATAATATTGCTCACAAAC 31
Db 1200 TTATTCAATAATATTGCTCACAAAC 1177

RESULT 10

US-08-831-310-8
;; Sequence 8, Application US/08831310
;; Publication No. US20020026035A1
;; GENERAL INFORMATION:
;; APPLICANT: Kleanthous, Harold et al.
;; TITLE OF INVENTION: Helicobacter GHPO 1360 and
;; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
;; NUMBER OF SEQUENCES: 18
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Clark & Elbing LLP
;; STREET: 176 Federal Street
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02110
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/831,310
;; FILING DATE: 01-APR-1997
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Clark, Paul T.
;; REGISTRATION NUMBER: 30,175
;; REFERENCE/DOCKET NUMBER: 06132/037001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-428-0200
;; TELEFAX: 617-428-7045
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 30 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-831-310-8
Query Match 69.7%; Score 21.6; DB 8; Length 30;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCTAATATTATTCAATAATATTGCTCAC 28
Db 3 GCTCAGTTATTCAATAATATTGCTCAC 30

RESULT 11
US-10-039-183A-8
;; Sequence 8, Application US/10039183A
;; Publication No. US20030143242A1
;; GENERAL INFORMATION:
;; APPLICANT: Kleanthous, Harold
;; APPLICANT: Lissolo, Ling
;; APPLICANT: Tomb, Jean-Francois
;; APPLICANT: Miller, Charles
;; APPLICANT: Al-Garawi, Amal
;; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
;; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
;; FILE REFERENCE: 06132/037002

; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-8

Query Match 69.7%; Score 21.6; DB 15; Length 30;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTTGCTCAC 28
Db 3 GCTCGATTTATTCATAATATTTGCTCAC 30

RESULT 12

US-10-425-115-28120
; Sequence 28120, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 28120
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRF4577_125658C.1
US-10-425-115-28120

Query Match 67.7%; Score 21; DB 18; Length 1290;
Best Local Similarity 82.8%; Pred. No. 9.7e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTTGCTCAC 29
Db 1107 GCCAATATTATTCATAATATTTGCTGATA 1135

RESULT 13

US-10-149-310-145
; Sequence 145, Application US/10149310
; Publication No. US20040077039A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; APPLICANT: Madden, Kevin T.
; APPLICANT: Sherman, Amir
; TITLE OF INVENTION: Modulation of Secondary Metabolite Production by
; FILE REFERENCE: 14184-019US1
; CURRENT APPLICATION NUMBER: US/10/149,310
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: PCT/US01/29288
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 60/233,564
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 145
; LENGTH: 2595
; TYPE: DNA
; ORGANISM: Kluyveromyces lactis
US-10-149-310-145

Query Match 67.7%; Score 21; DB 17; Length 2595;
Best Local Similarity 82.8%; Pred. No. 1.1e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCATAATATTTGCTCACAA 30
Db 1027 CTAATACTAAACAATAATATTTACTACTAA 1055

RESULT 14

US-10-149-310-147
; Sequence 147, Application US/10149310
; Publication No. US20040077039A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; APPLICANT: Madden, Kevin T.
; APPLICANT: Maxon, Mary
; APPLICANT: Sherman, Amir
; TITLE OF INVENTION: Modulation of Secondary Metabolite Production by
; FILE REFERENCE: 14184-019US1
; CURRENT APPLICATION NUMBER: US/10/149,310
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: PCT/US01/29288
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 60/233,564
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 147
; LENGTH: 2595
; TYPE: DNA
; ORGANISM: Kluyveromyces marxianus var. lactis
US-10-149-310-147

Query Match 67.7%; Score 21; DB 17; Length 2595;
Best Local Similarity 82.8%; Pred. No. 1.1e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCATAATATTTGCTCACAA 30
Db 1027 CTAATACTAAACAATAATATTTACTACTAA 1055

RESULT 15

US-10-473-126-189/c
; Sequence 189, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 189
; LENGTH: 5333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-189

Query Match 67.7%; Score 21; DB 18; Length 5333;
Best Local Similarity 82.8%; Pred. No. 1.2e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCAATAATATTGCTCACAA 30
|||||
Db 2983 CTAATATTATTAAATAATATTATTCTTAA 2955
|||||

Search completed: April 24, 2005, 05:52:22
Job time : 252.62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:31:54 ; Search time 1271.32 Seconds
(without alignments)
718.576 Million cell updates/sec

Title: US-10-039-183A-18

Perfect score: 24

Sequence: 1 cataacgcaataacgtacgcat 24

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gsa1:*

9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	80.8	663	8	AZ206030 SP_0106 B
2	18.8	78.3	486	2	BE011185 PM3-BN021
3	18.4	76.7	573	4	BI358245 RE44953.5
4	18.4	76.7	614	4	BI577150 RE70793.5
5	18.4	76.7	621	4	BI363229 RE47990.5
6	18.4	76.7	641	4	BI172492 RE15004.5
7	18.4	76.7	674	4	BI231884 RE24843.5
8	18.4	76.7	1201	9	AL057687 Drosophil
9	18.2	75.8	488	2	BE225571 MD0480 Me
10	18.2	75.8	731	8	CC111452 NDL.13J3.
11	17.8	74.2	617	9	AL737121 Danio rer
12	17.8	74.2	923	9	AL057860 Drosophil
13	17.6	73.3	533	1	AL915262 AL915262
14	17.6	73.3	646	7	CV105677 AGENCOURT
15	17.6	73.3	671	9	CG796914 ZMWBb034
16	17.6	73.3	731	9	EX132281 Danio rer
17	17.6	73.3	790	9	CC658434 OGREG31TV
18	17.6	73.3	875	9	CL496111 SAIL 61.D
19	17.4	72.5	816	9	CW006105 ZMWBLa000
20	17.4	72.5	945	9	CL985419 ZMWBHe000
21	17.4	72.5	1080	9	CW000366 ZMWBHf001
22	17.4	72.5	1100	9	CL998017 ZMWBHf001
23	17.2	71.7	182	2	BE707185 MR2-HT044
24	17.2	71.7	276	2	BB402258 BB402258

c	25	17.2	71.7	302	2	BB070814	BB070814
c	26	17.2	71.7	303	9	CG842150	Ynhw2764
c	27	17.2	71.7	430	7	CN629723	taf39a09.
c	28	17.2	71.7	464	5	EX609783	EX609783
c	29	17.2	71.7	556	7	CK498310	rawbb0.00
c	30	17.2	71.7	561	2	BF208786	BF208786
c	31	17.2	71.7	565	7	CN613438	601872561
c	32	17.2	71.7	585	7	CN629461	CN629461
c	33	17.2	71.7	599	6	CB039907	CB039907
c	34	17.2	71.7	601	1	AV400542	AV400542
c	35	17.2	71.7	609	5	BM389834	BM389834
c	36	17.2	71.7	617	5	BM097042	BM097042
c	37	17.2	71.7	619	5	BM183256	BM183256
c	38	17.2	71.7	620	1	AA543535	AA543535
c	39	17.2	71.7	622	5	BM138688	BM138688
c	40	17.2	71.7	626	5	BM134781	BM134781
c	41	17.2	71.7	629	5	BM138105	BM138105
c	42	17.2	71.7	665	5	BM096772	BM096772
c	43	17.2	71.7	668	6	CD868090	CD868090
c	44	17.2	71.7	686	8	BH932764	BH932764
c	45	17.2	71.7	689	8	BH946249	OBu92a03.

ALIGNMENTS

RESULT 1
AZ206030
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
ORIGIN

663 bp DNA linear GSS 31-AUG-2000
SP_0106_B1_F05_T7A Strongylocentrotus purpuratus, purple sea urchin, sperm Genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=106 Col=9 Row=L, genomic survey sequence.
AZ206030
GSS.
Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Ruchinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 663)
Cameron,R.A., Mahdrias,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingstone,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
10920195
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 106 row: L column: 9
Seq primer: T7
Class: BAC ends
High quality sequence stop: 663.
Location/Qualifiers
1. 663
/organism="Strongylocentrotus purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="Plate=106 Col=9 Row=L"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli DH10B"


```

/sex="male and female"
/dev stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

```

ORIGIN

```

Query Match          76.7%; Score 18.4; DB 4; Length 641;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 ATACGCAAAATACGCTACG 21

Db 272 ATACGCAAAATACGCTACG 291

RESULT 7

```

B1231884
LOCUS
DEFINITION
  RE24843.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE24843 5 similar to
  BG:DS01219.1; FBAN0004482 located on: 2L 35B9-35B9; 05/12/2001,
  mRNA sequence.

```

ACCESSION

B1231884

VERSION

B1231884.1 GI:14699148

KEYWORDS

EST.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 674)

AUTHORS

```

  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nuno,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.

```

BGDP/HMI RE Drosophila EST Project

Unpublished (2001)

CONTACT: Stapleton, M.

BGDP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_estofruitfly.berkeley.edu

hit genomic AB003645; arm:2r [14622887,14883853]

estimated-cyto:35B5-35B9; 05/12/2001

Plate: RE.248 row: D column: 7

High quality sequence stop: 634.

FEATURES

source

```

  1..674
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE24843"
  /sex="male and female"
  /dev stage="0-24 hours mixed stage embryonic"
  /lab_host="DH5-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pFlc-1"
  /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

```

ORIGIN

```

Query Match          76.7%; Score 18.4; DB 4; Length 674;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;

```

```

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ATACGCAAAATACGCTACG 21
  |||||
Db 272 ATACGCAAAATACGCTACG 291
  |||||

```

RESULT 8

```

CNS0014B/C
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR03F05 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.

```

ACCESSION

AL057687

VERSION

AL057687.1 GI:4931236

KEYWORDS

GSS.

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1201)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

  - Web : www.genoscope.cns.fr
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osogawa and
  Aaron Mammoss in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPCI-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain v2; cn bw sp, the same strain used for the BDGP's
  P1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

```

FEATURES

Location/Qualifiers

```

  1..1201
  /organism="Drosophila melanogaster"
  /mol_type="genomic DNA"
  /db_xref="taxon:7227"
  /clone="BACR03F05"
  /clone_lib="RPCI-98"
  /note="end : TET3"

```

ORIGIN

```

Query Match          76.7%; Score 18.4; DB 9; Length 1201;
Best Local Similarity 70.8%; Pred. No. 1.6e+02;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 CATACGCAAAATACGCTACGCTAT 24

Db 1142 CRKATCGCAAAATACGCTACGCTAT 1119

RESULT 9

```

BE225571/c
LOCUS
DEFINITION
  MD0480 Meloidogyne incognita J2 (#MD999-1) Meloidogyne incognita
  (Q19969), mRNA sequence.

```

ACCESSION

BE225571

VERSION

BE225571.1 GI:8930807

KEYWORDS

EST.

SOURCE

ORGANISM

Meloidogyne incognita (southern root-knot nematode)

Meloidogyne incognita

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;

Query Match 73.3%; Score 17.6; DB 7; Length 646;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CATACGCAATACGCTACGCAT 24
|||||
Db 409 CATAATGCAAAACCCCTACGCAT 386
|||||

RESULT 15
CG796914/c
LOCUS CG796914 671 bp DNA linear GSS 07-NOV-2003
DEFINITION ZMBBB0342E06.f ZMBBB Zea mays genomic clone ZMBBB0342E06 5',
genomic survey sequence.
ACCESSION CG796914
VERSION CG796914.1 GI:38215301
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 671)
Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.
and Wing, R.

TITLE Sequencing of the maize genome
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: <http://genome.arizona.edu>
PCR Primers

FORWARD: T7
BACKWARD: M13r
Plate: 0342 row: E column: 06
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..671
location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMBBB0342E06"
/lab_host="DH10B"
/clone_lib="ZMBBB"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII; Zea mays L. ssp. mays"

ORIGIN
Query Match 73.3%; Score 17.6; DB 9; Length 671;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CATACGCAATACGCTACGCAT 24
|||||
Db 317 CATAATGCAAAATACCCCTACGCAT 294
|||||

Search completed: April 24, 2005, 05:28:28
Job time : 1274.32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 02:04:39 ; Search time 194.028 Seconds
(without alignments)
750.949 Million cell updates/sec

Title: US-10-039-183A-18

Perfect score: 24

Sequence: 1 cataacgcaataacgtacgcat 24

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 303525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	8	US-08-831-310-18
2	24	100.0	24	15	US-10-039-183A-18
3	24	100.0	1000	9	US-09-881-752A-243
4	24	100.0	1149	8	US-08-831-310-1
5	24	100.0	1149	15	US-10-039-183A-1
6	19.2	80.0	456	17	US-10-335-977-481
7	19.2	80.0	900	17	US-10-335-977-483
8	18.4	76.7	1074	18	US-10-363-345A-39463
9	18.4	76.7	1074	18	US-10-363-345A-39464
10	18.4	76.7	1074	19	US-10-363-483A-39463
11	18.4	76.7	1074	19	US-10-363-483A-39464

c 12	18	75.0	996	18	US-10-363-345A-28743	Sequence 28743, A
c 13	18	75.0	996	18	US-10-363-345A-28744	Sequence 28744, A
c 14	18	75.0	996	19	US-10-363-483A-28743	Sequence 28743, A
c 15	18	75.0	996	19	US-10-363-483A-28744	Sequence 28744, A
c 16	17.6	73.3	261	18	US-10-425-115-22981	Sequence 22981, A
c 17	17.6	73.3	719	15	US-10-311-455-2374	Sequence 2374, A
c 18	17.6	73.3	1757	15	US-10-311-455-2394	Sequence 2394, A
c 19	17.4	72.5	502	18	US-10-363-345A-12779	Sequence 12779, A
c 20	17.4	72.5	502	18	US-10-363-345A-12780	Sequence 12780, A
c 21	17.4	72.5	502	19	US-10-363-483A-12779	Sequence 12779, A
c 22	17.4	72.5	502	19	US-10-363-483A-12780	Sequence 12780, A
c 23	17.2	71.7	575	18	US-10-363-345A-37981	Sequence 37981, A
c 24	17.2	71.7	575	18	US-10-363-345A-37982	Sequence 37982, A
c 25	17.2	71.7	575	19	US-10-363-483A-37981	Sequence 37981, A
c 26	17.2	71.7	575	19	US-10-363-483A-37982	Sequence 37982, A
c 27	17.2	71.7	638	18	US-10-363-345A-8061	Sequence 8061, A
c 28	17.2	71.7	638	18	US-10-363-345A-8062	Sequence 8062, A
c 29	17.2	71.7	638	19	US-10-363-483A-8061	Sequence 8061, A
c 30	17.2	71.7	638	19	US-10-363-483A-8062	Sequence 8062, A
c 31	17.2	71.7	768	18	US-10-363-345A-6289	Sequence 6289, A
c 32	17.2	71.7	768	18	US-10-363-345A-6290	Sequence 6290, A
c 33	17.2	71.7	768	19	US-10-363-483A-6289	Sequence 6289, A
c 34	17.2	71.7	768	19	US-10-363-483A-6290	Sequence 6290, A
c 35	17.2	71.7	1386	17	US-10-282-122A-41866	Sequence 41866, A
c 36	16.8	70.0	536	18	US-10-363-345A-40021	Sequence 40021, A
c 37	16.8	70.0	536	18	US-10-363-345A-40022	Sequence 40022, A
c 38	16.8	70.0	536	19	US-10-363-483A-40021	Sequence 40021, A
c 39	16.8	70.0	536	19	US-10-363-483A-40022	Sequence 40022, A
c 40	16.8	70.0	716	18	US-10-437-963-49304	Sequence 49304, A
c 41	16.8	70.0	869	18	US-10-363-345A-35875	Sequence 35875, A
c 42	16.8	70.0	869	18	US-10-363-345A-35876	Sequence 35876, A
c 43	16.8	70.0	869	19	US-10-363-483A-35875	Sequence 35875, A
c 44	16.8	70.0	869	19	US-10-363-483A-35876	Sequence 35876, A
c 45	16.8	70.0	1330	18	US-10-363-345A-35471	Sequence 35471, A

ALIGNMENTS

RESULT 1

US-08-831-310-18
; Sequence 18, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPD 1360 and
; TITLE OF INVENTION: GHPD 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831.310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-831-310-18
Query Match 100.0%; Score 24; DB 8; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.41; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATAACGCAAAATAACGCTACGCAT 24
   |||||
Db 1 CATAACGCAAAATAACGCTACGCAT 24

RESULT 2
US-10-039-183A-18
; Sequence 18, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; FILE REFERENCE: 06132/037002
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 24
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-18

Query Match 100.0%; Score 24; DB 15; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.41; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATAACGCAAAATAACGCTACGCAT 24
   |||||
Db 1 CATAACGCAAAATAACGCTACGCAT 24

RESULT 3
US-09-881-752A-243
; Sequence 243, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 243
; LENGTH: 1000
; TYPE: DNA
```

```
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (51)...(947)
US-09-881-752A-243

Query Match 100.0%; Score 24; DB 9; Length 1000;
Best Local Similarity 100.0%; Pred. No. 0.61; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATAACGCAAAATAACGCTACGCAT 24
   |||||
Db 123 CATAACGCAAAATAACGCTACGCAT 146

RESULT 4
US-08-831-310-1
; Sequence 1, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 106...1002
; OTHER INFORMATION:
; NAME/KEY: Signal Sequence
; LOCATION: 106...166
; OTHER INFORMATION:
US-08-831-310-1

Query Match 100.0%; Score 24; DB 8; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0.62; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATAACGCAAAATAACGCTACGCAT 24
   |||||
Db 178 CATAACGCAAAATAACGCTACGCAT 201
```

RESULT 5

US-10-039-183A-1
; Sequence 1, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Liessolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (106)...(1002)
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: (106)...(166)
US-10-039-183A-1

Query Match 100.0%; Score 24; DB 15; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATAACGCAATAACGCTACGCAT 24
|||
Db 178 CATAACGCAATAACGCTACGCAT 201

RESULT 6

US-10-335-977-481
; Sequence 481, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; City: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 481:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...456
; SEQUENCE DESCRIPTION: SEQ ID NO: 481:
US-10-335-977-481

Query Match 80.0%; Score 19.2; DB 17; Length 456;
Best Local Similarity 87.5%; Pred. No. .79;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CATAACGCAATAACGCTACGCAT 24
|||
Db 88 CATAATGCGAATAACTCTACGCAT 111

RESULT 7
US-10-335-977-483
; Sequence 483, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; City: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 483:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

```
/ ORGANISM: Helicobacter pylori
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...900
/ SEQUENCE DESCRIPTION: SEQ ID NO: 483:
US-10-333-977-483

Query Match      80.0%; Score 19.2; DB 17; Length 900;
Best Local Similarity 87.5%; Pred. No. 85;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATAACGCAATAACGCTACGCAT 24
Db 73 CATAATGCGAATAACTCTACGCAT 96

RESULT 8
US-10-363-345A-39463/c
; Sequence 39463, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 39463
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 39463
; NAME/KEY: unsure
; LOCATION: (883, 910, 912, 922..923, 925, 931..932, 934, 939..940, 949..)
US-10-363-345A-39463

Query Match      76.7%; Score 18.4; DB 18; Length 1074;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TAACGCAATAACGCTACGC 22
Db 875 TAACGCGAATAACGCTACGC 856

RESULT 9
US-10-363-345A-39464
; Sequence 39464, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; TITLE OF INVENTION: cytosines in genomic DNA in the sequence context of 5'-CpG-3
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 39464
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 39464
; NAME/KEY: unsure
```

```
/ NAME/KEY: unsure
/ LOCATION: (124..125, 135..136, 141..142, 144, 150..151, 153, 163, 165)
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (192)
US-10-363-345A-39464

Query Match      76.7%; Score 18.4; DB 18; Length 1074;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TAACGCAATAACGCTACGC 22
Db 200 TAACGCGAATAACGCTACGC 219

RESULT 10
US-10-363-483A-39463/c
; Sequence 39463, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 39463
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 39463
; NAME/KEY: unsure
; LOCATION: (883, 910, 912, 922..923, 925, 931..932, 934, 939..940, 949..)
US-10-363-483A-39463

Query Match      76.7%; Score 18.4; DB 19; Length 1074;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TAACGCAATAACGCTACGC 22
Db 875 TAACGCGAATAACGCTACGC 856

RESULT 11
US-10-363-483A-39464
; Sequence 39464, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; TITLE OF INVENTION: illnesses
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 39464
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 39464
; NAME/KEY: unsure
```

```
; NAME/KEY: unsure
; LOCATION: (124...125, 135...136, 141...142, 144, 150...151, 153, 163, 165)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (192)
US-10-363-483A-39464

Query Match      76.7%; Score 18.4; DB 19; Length 1074;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TAACGCAATAACGCTACGC 22
   |||||
Db 200 TAACGCAATAACGCTACGC 219

RESULT 12
US-10-363-345A-28743/C
; Sequence 28743, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 28743
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 28743
US-10-363-345A-28743

Query Match      75.0%; Score 18; DB 18; Length 996;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ACGCAATAACGCTACGC 22
   |||||
Db 369 ACGCAATAACGCTACGC 352

RESULT 13
US-10-363-345A-28744
; Sequence 28744, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 28744
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 28744
US-10-363-345A-28744

Query Match      75.0%; Score 18; DB 18; Length 996;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ACGCAATAACGCTACGC 22
   |||||
Db 628 ACGCAATAACGCTACGC 645

RESULT 14
US-10-363-483A-28743/C
; Sequence 28743, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 28743
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 28743
US-10-363-483A-28743

Query Match      75.0%; Score 18; DB 19; Length 996;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ACGCAATAACGCTACGC 22
   |||||
Db 369 ACGCAATAACGCTACGC 352

RESULT 15
US-10-363-483A-28744
; Sequence 28744, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 28744
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: Cpg-island No: 28744
US-10-363-483A-28744

Query Match      75.0%; Score 18; DB 19; Length 996;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ACGCAATAACGCTACGC 22
   |||||
Db 628 ACGCAATAACGCTACGC 645

Search completed: April 24, 2005, 05:52:23
Job time : 195.028 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:51:24 ; Search time 46.6479 Seconds
(without alignments)
841.853 Million cell updates/sec

Title: US-10-039-183A-18
Perfect score: 24
Sequence: 1 cataacgcaataacgctacgcacat 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A-COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B-COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A-COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B-COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCUTUS-COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	1149	US-09-336-115C-23	Sequence 23, Appl
2	16.8	70.0	1209	US-09-540-236-315	Sequence 315, Appl
3	16.8	70.0	100848	US-09-596-002-39	Sequence 39, Appl
4	16.6	69.2	601	US-09-949-016-203369	Sequence 203369, Appl
5	16.6	69.2	76962	US-09-949-016-17482	Sequence 17482, A
6	16.4	68.3	7886	US-09-453-702B-102	Sequence 102, Appl
7	16.2	67.5	1569	US-09-107-532A-2563	Sequence 2563, Appl
8	16.2	67.5	1977	US-09-602-777A-395	Sequence 395, Appl
9	16.2	67.5	120213	US-09-949-016-13304	Sequence 13304, A
10	16.2	67.5	120217	US-09-949-016-12260	Sequence 12260, A
11	16	66.7	169	US-09-270-767-2455	Sequence 2455, Appl
12	16	66.7	169	US-09-270-767-17737	Sequence 17737, A
13	16	66.7	1023	US-09-328-352-2578	Sequence 2578, Appl
14	16	66.7	100836	US-09-949-016-12871	Sequence 12871, A
15	16	66.7	100837	US-09-949-016-17063	Sequence 17063, A
16	16	66.7	147321	US-09-949-016-15450	Sequence 15450, A
17	15.6	65.0	1040	US-08-936-165A-229	Sequence 229, Appl
18	15.6	65.0	1167	US-09-489-039A-2606	Sequence 2606, Appl
19	15.6	65.0	2280	US-09-328-352-3524	Sequence 3524, Appl
20	15.6	65.0	2475	US-09-489-039A-310	Sequence 310, Appl
21	15.6	65.0	5614	US-08-956-171E-99	Sequence 99, Appl
22	15.6	65.0	5614	US-08-781-986A-99	Sequence 99, Appl
23	15.6	65.0	43360	US-09-453-702B-206	Sequence 206, Appl
24	15.6	65.0	43325	US-09-453-702B-261	Sequence 261, Appl
25	15.6	65.0	1830121	US-09-557-884-1	Sequence 1, Appl
26	15.6	65.0	1830121	US-09-643-990A-1	Sequence 1, Appl
27	15.4	64.2	1851	US-09-248-796A-889	Sequence 889, Appl

c 28	15.2	63.3	402	4	US-09-634-238-96	Sequence 96, Appl
c 29	15.2	63.3	498	3	US-09-134-001C-774	Sequence 774, Appl
c 30	15.2	63.3	561	4	US-09-489-039A-5236	Sequence 5236, Appl
c 31	15.2	63.3	601	4	US-09-949-016-44723	Sequence 44723, A
c 32	15.2	63.3	2332	1	US-08-252-492-1	Sequence 1, Appl
c 33	15.2	63.3	2332	2	US-08-727-126-1	Sequence 1, Appl
c 34	15.2	63.3	2332	3	US-08-942-761-1	Sequence 1, Appl
c 35	15.2	63.3	2934	3	US-09-206-942-52	Sequence 52, Appl
c 36	15.2	63.3	2952	3	US-09-206-942-50	Sequence 50, Appl
c 37	15.2	63.3	3015	3	US-09-206-942-56	Sequence 56, Appl
c 38	15.2	63.3	3033	3	US-09-206-942-54	Sequence 54, Appl
c 39	15.2	63.3	3131	4	US-09-710-279-4137	Sequence 4137, Appl
c 40	15.2	63.3	3966	4	US-09-978-594-555	Sequence 555, Appl
c 41	15.2	63.3	5838	4	US-08-956-171E-66	Sequence 66, Appl
c 42	15.2	63.3	5838	4	US-08-781-986A-66	Sequence 66, Appl
c 43	15.2	63.3	41132	4	US-09-949-016-12021	Sequence 12021, A
c 44	15.2	63.3	41136	4	US-09-949-016-13021	Sequence 13021, A
c 45	15.2	63.3	70770	4	US-09-949-016-16938	Sequence 16938, A

ALIGNMENTS

RESULT 1
US-09-336-115C-23
; Sequence 23, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (106)...(1002)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (106)...(166)
US-09-336-115C-23

Query Match 100.0% Score 24; DB 4; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATAACGCAATAACGCTACGCAT 24
Db 178 CATAACGCAATAACGCTACGCAT 201

RESULT 2
US-09-540-236-315/c
; Sequence 315, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Braton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATA
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 315

; LENGTH: 1209
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-315

Query Match 70.0%; Score 16.8; DB 4; Length 1209;
Best Local Similarity 90.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATAACGCAATAACGCTAC 20
|||||
Db 935 CATAACGCAATAATGCTTC 916

RESULT 3

US-09-596-002-39/c
; Sequence 39, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 100848
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 39
; PUBLICATION INFORMATION:
US-09-596-002-39

Query Match 70.0%; Score 16.8; DB 4; Length 100848;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CATAACGCAATAACGCTAC 20
|||||
Db 35547 CATAACGCAATAATGCTTC 35528

RESULT 4

US-09-949-016-203369
; Sequence 203369, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203369
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-203369

Query Match 69.2%; Score 16.6; DB 4; Length 601;
Best Local Similarity 82.6%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CATAACGCAATAACGCTACGCA 23
|||||
Db 557 CATAACCCAGATACTCTACTCA 579

RESULT 5

US-09-949-016-17482/c
; Sequence 17482, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17482
; LENGTH: 76962
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(76962)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17482

Query Match 69.2%; Score 16.6; DB 4; Length 76962;
Best Local Similarity 82.6%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CATAACGCAATAACGCTACGCA 23
|||||
Db 13226 CATAACCCAGATACTCTACTCA 13204

RESULT 6

US-09-453-702B-102/c
; Sequence 102, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Berna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27396
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 7886
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-453-702B-102

Query Match 68.3%; Score 16.4; DB 3; Length 7886;
Best Local Similarity 94.4%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CATAACGCAATAACGCT 18
Db 1083 CATAACGCAATAACGCT 1066

RESULT 7
US-09-107-532A-2563
Sequence 2563, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2563:
SEQUENCE CHARACTERISTICS:
LENGTH: 1569 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1569
SEQUENCE DESCRIPTION: SEQ ID NO: 2563:
US-09-107-532A-2563

Query Match 67.5%; Score 16.2; DB 4; Length 1569;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AACGCAATAACGCTACGCAT 24
Db 721 AACTCACTAAGATACGCAT 741

RESULT 8
US-09-602-777A-395/c
Sequence 395, Application US/09602777A
Patent No. 6831165
GENERAL INFORMATION:
APPLICANT: Pompejus, Markus
APPLICANT: Kroger, Burkhard
APPLICANT: Schroder, Hartwig
APPLICANT: Zelder, Oskar
APPLICANT: Haberhauer, Gregor
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN HOMEOSTASIS AND ADAPTATION
FILE REFERENCE: BGI-128CP
CURRENT APPLICATION NUMBER: US/09/602,777A
CURRENT FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/141031
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: DE 19931636.8
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19932125.6
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932126.4
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932127.2
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932128.0
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932129.9
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: DE 19932226.0
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: DE 19932920.6
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19932922.2
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19932924.9
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19932928.1
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19932930.3
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19932933.8
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19932935.4
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19932973.7
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19933002.6
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19933003.4
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19933005.0

; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933006.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19941378.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941390.8
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941391.6
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19942088.2
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 395
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1954)
; OTHER INFORMATION: RXN02508
US-09-602-777A-395

Query Match 67.5%; Score 16.2; DB 4; Length 1977;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 AACGCAATAAAGCTACGCAT 24
||||| ||||| ||||| |||||
Db 887 AACGCAAGAACGCAACCCAT 867

RESULT 9

US-09-949-016-13304/c
; Sequence 13304, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13304
; LENGTH: 120213
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(120213)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13304

Query Match 67.5%; Score 16.2; DB 4; Length 120213;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATAACGCAATAAAGCTACG 21
||||| ||||| ||||| |||||
Db 97497 CATAACACAATAATGCTAAG 97477

RESULT 10

US-09-949-016-12260/c

; Sequence 12260, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12260
; LENGTH: 120217
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(120217)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12260

Query Match 67.5%; Score 16.2; DB 4; Length 120217;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATAACGCAATAAAGCTACG 21
||||| ||||| ||||| |||||
Db 97501 CATAACACAATAATGCTAAG 97481

RESULT 11

US-09-270-767-2455/c
; Sequence 2455, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2455
; LENGTH: 169
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-2455

Query Match 66.7%; Score 16; DB 4; Length 169;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CATAACGCAATAAAGCTACGCAT 24
||||| ||||| ||||| |||||
Db 50 CTTAACGCCCATAGGCTAAGCAT 27

RESULT 12

US-09-270-767-17737/c
; Sequence 17737, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 10:03:10 ; Search time 157.69 Seconds
(without alignments)
900.968 Million cell updates/sec

Title: US-10-039-183A-18

Perfect score: 24
Sequence: 1 cataacgcaataacgctacgcat 24

Scoring table: IDENTITY NUC

Gapop 10'0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	AAV07973	Helicobac
2	24	100.0	1000	Aax14041	H. pylori
3	24	100.0	1149	AAV07963	Helicobac
4	24	100.0	1149	AAD61576	Helicobac
5	22.4	93.3	1082	AAV90844	Nucleotid
6	22.4	93.3	1082	AAV90591	Nucleotid
7	22.4	93.3	1117	AAV90580	Nucleotid
8	19.2	80.0	456	AAV24638	H. pylori
9	19.2	80.0	900	AAV24859	H. pylori
10	18.4	76.7	1074	ABQ52872	Oligonucl
11	18.4	76.7	1074	ABQ52873	Oligonucl
12	18	75.0	996	ABQ42152	Oligonucl
13	18	75.0	996	ABQ42153	Oligonucl
14	17.6	73.3	719	ABL34401	Human imm
15	17.6	73.3	1757	ABL34421	Human imm
16	17.4	72.5	502	ABQ26188	Oligonucl
17	17.4	72.5	502	ABQ26189	Oligonucl
18	17.4	72.5	7309	ABL29678	Drosophil
19	17.2	71.7	575	ABQ51390	Oligonucl
20	17.2	71.7	575	ABQ51391	Oligonucl

21	17.2	71.7	638	6	ABQ21471	Oligonucl
22	17.2	71.7	638	6	ABQ21470	Oligonucl
23	17.2	71.7	768	6	ABQ19698	Oligonucl
24	17.2	71.7	768	6	ABQ19699	Oligonucl
25	17.2	71.7	1386	8	ACA53996	Prokaryot
26	17.2	71.7	110000	2	AAZ01425_06	Continuation (7 of
27	16.8	70.0	536	6	ABQ53430	Oligonucl
28	16.8	70.0	536	6	ABQ53431	Oligonucl
29	16.8	70.0	869	6	ABQ49284	Oligonucl
30	16.8	70.0	869	6	ABQ49285	Oligonucl
31	16.8	70.0	1209	12	ADL02629	DNA encod
32	16.8	70.0	1330	6	ABQ48881	Oligonucl
33	16.8	70.0	1330	6	ABQ48880	Oligonucl
34	16.8	70.0	6112	6	ABK33980	Human DNA
35	16.8	70.0	6112	8	ADA20368	Prostate
36	16.8	70.0	6112	8	ADA84175	Human ren
37	16.8	70.0	100848	4	AAF28552	Genomic f
38	16.6	69.2	541	6	ABQ21866	Oligonucl
39	16.6	69.2	541	6	ABQ21867	Oligonucl
40	16.6	69.2	563	6	ABQ46821	Oligonucl
41	16.6	69.2	563	6	ABQ46820	Oligonucl
42	16.6	69.2	584	6	ABQ19238	Oligonucl
43	16.6	69.2	584	6	ABQ19239	Oligonucl
44	16.6	69.2	587	6	ABQ51058	Oligonucl
45	16.6	69.2	587	6	ABQ51059	Oligonucl

ALIGNMENTS

RESULT 1

AAV07973

ID AAV07973 standard; DNA; 24 BP.

XX AAV07973;

AC AAV07973;

DT 25-MAR-2003 (revised)

DT 02-FEB-1999 (first entry)

XX Helicobacter pylori polypeptide GHPO 1360 3' DNA primer.

XX GHPO 1360; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;

KW primer; ss.

XX Synthetic.

OS Helicobacter pylori.

XX MO9843479-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

XX 01-APR-1997; 97US-00831310.

XX 01-APR-1997; 97US-00834666.

PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.

XX New isolated Helicobacter polynucleotides - used to develop products for

PT the diagnosis, prevention and treatment of Helicobacter infections and

PT gastroduodenal diseases.

PS Claim 5; Page 156; 184pp; English.

XX This 3' primer is used with a 5' primer (see AAV07969) in the PCR

CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic DNA in

CC order to obtain DNA (see AAV07963) encoding the unprocessed form of a 32

CC kDa polypeptide (see AAW73034) designated GHPO 1360. The 3' primer was

Query Match 100.0%; Score 24; DB 2; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATACGCAAAATAACGCTACGCAT 24
 |||||
 Db 178 CATACGCAAAATAACGCTACGCAT 201
 |||||

RESULT 4
 AAD61576
 ID AAD61576 standard; DNA; 1149 BP.
 XX AC
 XX AAD61576;
 XX 15-JAN-2004 (first entry)
 XX DE Helicobacter pylori p32 antigen DNA.
 XX KW Immune response; Helicobacter infection; adjuvant; heat-labile toxin; LT;
 KW cholera toxin; CT; urease; therapy; antibacterial; vaccine; antigen; p32;
 KW gene; ds.
 XX OS Helicobacter pylori.
 XX FH Key Location/Qualifiers
 FT CDS 106..1005
 FT /tag= a
 FT /product= "Antigen p32"
 FT sig_peptide 106..166
 FT /tag= b
 FT mat_peptide 167..1002
 FT /tag= c
 FT /product= "Mature p32"
 XX US6576244-B1.
 XX 10-JUN-2003.
 XX 18-JUN-1999; 99US-00336115.
 XX 19-JUN-1998; 98US-00100258.
 XX (ACAM-) ACAMBIS INC.
 XX PI Weltzin RA, Guy B;
 XX WPI; 2003-799824/75.
 XX P-PSDB; ABW00787.
 XX Inducing immune response to Helicobacter useful for treating Helicobacter
 PT pylori infection, by administering immunogenic Helicobacter polypeptide
 PT admixed with adjuvant having heat-labile toxin of Escherichia coli.
 XX Claim 8; Col 105-108; Opp; English.
 XX The invention relates to a method for inducing protective immune response
 CC to Helicobacter infection. The method comprising administering to a
 CC mammal by injection an immunogenic Helicobacter pylori polypeptide
 CC comprising a subunit of H.pylori urease admixed with an adjuvant having
 CC one or more heat-labile toxin of Escherichia coli (LT), B subunit of LT
 CC (LTB), cholera toxin (CT), and B subunit of CT. The method is useful for
 CC inducing an immune response to Helicobacter infection in a mammal. The
 CC method is useful for both treatment and prevention of H.pylori infection.
 CC The invention is useful as vaccine. The present sequence is Helicobacter
 CC pylori p32 antigen DNA
 XX SQ Sequence 1149 BP; 412 A; 185 C; 242 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 24; DB 10; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATACGCAAAATAACGCTACGCAT 24
 |||||
 Db 178 CATACGCAAAATAACGCTACGCAT 201
 |||||

RESULT 5
 AAV90844
 ID AAV90844 standard; DNA; 1082 BP.
 XX AC AAV90844;
 XX 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX DE Nucleotide sequence of cluster 3.
 XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 XX OS Helicobacter pylori.
 XX PN WO9849314-A2.
 XX 05-NOV-1998.
 XX 25-APR-1998; 98WO-US008487.
 XX 25-APR-1997; 97US-0045107P.
 XX 14-OCT-1997; 97US-0061958P.
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 XX PI Chow TP, Fry KE, Lim MY, Mcatee CP;
 XX WPI; 1999-009433/01.
 XX New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX Claim 20; Page 281-282; 402pp; English.
 XX The present sequence encodes a Helicobacter pylori antigenic protein that
 CC is characterised by immunoreactivity with H. pylori-positive antisera.
 CC The proteins are highly immunogenic and induce a long-lasting immune
 CC response that persists even after antimicrobial treatment. In antibody-
 CC detection assays, on sera, plasma, urine, saliva etc., they are highly
 CC sensitive and specific. The specification also describes 69 previously
 CC unrecognised immunogenic cluster families. H. pylori antigens are used to
 CC detect H. pylori-specific antibodies, for diagnosing infection or to
 CC confirm eradication of infection, and in vaccines to protect against H.
 CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
 CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
 XX SQ Sequence 1082 BP; 383 A; 188 C; 222 G; 286 T; 0 U; 3 Other;

Query Match 93.3%; Score 22.4; DB 2; Length 1082;
 Best Local Similarity 95.8%; Pred. No. 1.1;
 Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CATACGCAAAATAACGCTACGCAT 24
 |||||
 Db 237 CATACGCAAGATACGCTACGCAT 260
 |||||

RESULT 6
 AAV90591
 ID AAV90591 standard; DNA; 1082 BP.
 XX AC AAV90591;
 XX 20-MAR-2003 (revised)


```

PR 06-DEC-1996; 96US-00761318.
XX (ASTR ) ASTRA AB.
PA Smith D, Alm RA;
PI WPI; 1997-503122/46.
XX P-PSDB; AAW55229.
XX
XX Helicobacter pylori nucleic acid sequences and encoded polypeptide(s) -
PT useful in vaccines to treat or prevent H. pylori infection and for
PT diagnosis of H. pylori infection.
XX
PS Claim 5,6; Page 155; 1145pp; English.
XX
XX This sequence encodes a Helicobacter pylori protein of unspecified
CC function. The protein may be used in a vaccine to prevent or treat H.
CC pylori infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of H.
CC pylori in a sample, and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences, and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
XX hosts
XX
SQ Sequence 456 BP; 170 A; 79 C; 106 G; 101 T; 0 U; 0 Other;
Query Match 80.0%; Score 19.2; DB 2; Length 456;
Best Local Similarity 87.5%; Pred. No. 33;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATAACGCAATACGCTACGCAT 24
DB 88 CATAATGCGAATAACTCTACGCAT 111
||||| ||||| ||||| ||||| |||||
||||| ||||| ||||| ||||| |||||

RESULT 9
AAV24859
ID AAV24859 standard; DNA; 900 BP.
XX
XX AAV24859;
AC
DT 24-JUN-1998 (first entry)
XX
DE H. pylori cell envelope OMP ORF 02ae11612_22477267_f2_27.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
XX
OS Helicobacter pylori.
XX
XX Key Location/Qualifiers
XX CDS 1..900
XX /*tag= a
XX
XX WO9737044-A1.
XX
XX 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US005223.
XX

```

```

PR 29-MAR-1996; 96US-00625811.
PR 02-APR-1996; 96US-00758731.
PR 25-OCT-1996; 96US-00736905.
PR 28-OCT-1996; 96US-00738859.
PR 06-DEC-1996; 96US-00761318.
XX (ASTR ) ASTRA AB.
PA Smith D, Alm RA;
PI WPI; 1997-503122/46.
XX P-PSDB; AAW55450.
XX
XX Helicobacter pylori nucleic acid sequences and encoded polypeptide(s) -
PT useful in vaccines to treat or prevent H. pylori infection and for
PT diagnosis of H. pylori infection.
XX
PS Claim 5,6,21; Page 288; 1145pp; English.
XX
XX This sequence encodes a H. pylori cell envelope outer membrane protein
CC (OMP) having no terminal Phe residue. The protein may be used in a
CC vaccine to prevent or treat H. pylori infection or to identify H. pylori
CC polypeptide binding compounds, useful as potential H. pylori life cycle
CC activators or inhibitors. The DNA and probes derived from it may be used
CC for the identification of H. pylori in a sample and the diagnosis of H.
CC pylori infection. Nucleic acid sequences complementary to the DNA act as
CC antisense sequences and can be used to prevent the translation of H.
CC pylori mRNA. Antibodies against the protein can be used in immunoassays
CC to evaluate the abundance and distribution of H. pylori-specific
CC antigens. The genomic sequence of H. pylori (ATCC 55679) was determined
CC from overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides, and
CC the predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts
XX
SQ Sequence 900 BP; 329 A; 161 C; 193 G; 217 T; 0 U; 0 Other;
Query Match 80.0%; Score 19.2; DB 2; Length 900;
Best Local Similarity 87.5%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CATAACGCAATACGCTACGCAT 24
DB 73 CATAATGCGAATAACTCTACGCAT 96
||||| ||||| ||||| ||||| |||||
||||| ||||| ||||| ||||| |||||

RESULT 10
ABQ52872/c
ID ABQ52872 standard; DNA; 1074 BP.
XX
XX ABQ52872;
AC
DT 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 39463.
DE
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX

```

```

XX 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX (EPIG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1074 BP; 175 A; 114 C; 375 G; 394 T; 0 U; 16 Other;
Query Match 76.7%; Score 18.4; DB 6; Length 1074;
Best Local Similarity 95.0%; Pred. No. 81;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 TAACGCAATAACGCTACGC 22
Db 875 TAACGCGAATAACGCTACGC 856
RESULT 11
ABQ52873
ID ABQ52873 standard; DNA; 1074 BP.
XX
XX ABQ52873;
AC
XX
XX 12-JUL-2002 (first entry)
DT
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 39464.
DE
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX WO200218632-A2.
FN
XX
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP010074.
PF
XX
XX 01-SEP-2000; 2000DE-01043826.
PR
XX 05-SEP-2000; 2000DE-01044543.
PR
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX (EPIG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1074 BP; 175 A; 114 C; 375 G; 394 T; 0 U; 16 Other;
Query Match 76.7%; Score 18.4; DB 6; Length 1074;
Best Local Similarity 95.0%; Pred. No. 81;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 TAACGCAATAACGCTACGC 22
Db 875 TAACGCGAATAACGCTACGC 856
RESULT 11
ABQ52873
ID ABQ52873 standard; DNA; 1074 BP.
XX
XX ABQ52873;
AC
XX
XX 12-JUL-2002 (first entry)
DT
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 39464.
DE
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX WO200218632-A2.
FN
XX
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP010074.
PF
XX
XX 01-SEP-2000; 2000DE-01043826.
PR
XX 05-SEP-2000; 2000DE-01044543.
PR
XX (EPIG-) EPIGENOMICS AG.
PA

```

PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX
 XX Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 PS
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 XX SQ Sequence 996 BP; 130 A; 113 C; 352 G; 401 T; 0 U; 0 Other;
 Query Match 75.0%; Score 18; DB 6; Length 996;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 5 ACGCAATAACGCTACGC 22
 DB 369 ACGCAATAACGCTACGC 352
 RESULT 13
 ABQ42153
 ID ABQ42153 standard; DNA; 996 BP.
 XX
 XX AC ABQ42153;
 XX
 XX DT 12-JUL-2002 (first entry)
 XX
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 28744.
 XX
 XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 XX OS Homo sapiens.
 XX
 XX PN WO200218632-A2.
 XX
 XX PD 07-MAR-2002.
 XX
 XX PP 01-SEP-2001; 2001WO-EP010074.
 XX
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX PR 05-SEP-2000; 2000DE-01044543.
 XX
 XX PA (EPIG-) EPIGENOMICS AG.
 XX
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX

XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 PS
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 XX SQ Sequence 996 BP; 401 A; 352 C; 113 G; 130 T; 0 U; 0 Other;
 Query Match 75.0%; Score 18; DB 6; Length 996;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 5 ACGCAATAACGCTACGC 22
 DB 628 ACGCAATAACGCTACGC 645
 RESULT 14
 ABL34401/C
 ID ABL34401 standard; DNA; 719 BP.
 XX
 XX AC ABL34401;
 XX
 XX DT 26-MAR-2002 (first entry)
 XX
 XX DE Human immune system associated gene SEQ ID NO: 2374.
 XX
 XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 XX
 XX OS Homo sapiens.
 XX
 XX PN WO200200928-A2.
 XX
 XX PD 03-JAN-2002.
 XX
 XX PP 02-JUL-2001; 2001WO-EP007537.
 XX
 XX PR 30-JUN-2000; 2000DE-01032529.
 XX
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX PA (EPIG-) EPIGENOMICS AG.
 XX
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-371829/40.
 XX

DR WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 2374; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 719 BP; 116 A; 76 C; 194 G; 333 T; 0 U; 0 Other;
Query Match 73.3%; Score 17.6; DB 6; Length 719;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CATAACGCAAAATAACGCTACGCAT 24
DB 73 CAATAACGCAAAATAACGCTACGCAT 50
RESULT 15
ID ABL34421/c
XX ABL34421 standard; DNA; 1757 BP.
XX
AC ABL34421;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2394.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 2394; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 719 BP; 116 A; 76 C; 194 G; 333 T; 0 U; 0 Other;
Query Match 73.3%; Score 17.6; DB 6; Length 719;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CATAACGCAAAATAACGCTACGCAT 24
DB 73 CAATAACGCAAAATAACGCTACGCAT 50

CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 1757 BP; 396 A; 83 C; 464 G; 813 T; 0 U; 1 Other;
Query Match 73.3%; Score 17.6; DB 6; Length 1757;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CATAACGCAAAATAACGCTACGCAT 24
DB 280 CGTAACCTAAATAACGATACGCAT 257
Search completed: April 24, 2005, 02:20:10
Job time : 159.69 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 17:27:10 ; Search time 622.986 Seconds
(without alignments)
1866.696 Million cell updates/sec

Title: US-10-039-183A-18

Perfect score: 24
Sequence: 1 cataacgcaataacgctacscat 24

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_ats.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	24	6	BD082356
2	24	100.0	1000	6	BD092341
3	24	100.0	1149	6	AR342398 Sequence
4	24	100.0	1149	6	BD082346
5	24	100.0	11421	1	AE000538 Helicobac
6	22.4	93.3	1082	6	BD061716 Antigenic
7	22.4	93.3	1082	6	BD061705 Antigenic
8	22.4	93.3	1117	6	BD061705 Antigenic
9	19.2	80.0	10085	1	AE001454 Helicobac
10	18.2	75.8	11193	1	AE013426 Methanosa
11	18.2	75.8	128403	2	AC017472 Drosophil
12	18.2	75.8	152607	3	AC010922 Drosophil
13	18.2	75.8	163466	3	AC010921 Drosophil
14	18.2	75.8	298641	3	AE003504 Drosophil
15	17.8	74.2	37578	3	LMPL171 Leishmani
16	17.8	74.2	44670	2	AC019551 Drosophil
17	17.8	74.2	89464	2	AC138543 Magnapor
18	17.8	74.2	116018	2	AC127424 Magnapor
19	17.8	74.2	163553	5	BX640403 Zebrafish

C 20	17.8	74.2	186686	3	AC008192	Drosophil
C 21	17.8	74.2	212555	2	BX640460	Danio rer
C 22	17.8	74.2	247876	3	AE003726	Drosophil
C 23	17.6	73.3	264	8	AY021665	Oryza sat
C 24	17.6	73.3	719	6	AX347303	Sequence
C 25	17.6	73.3	1757	6	AX347323	Sequence
C 26	17.6	73.3	2813	1	AB119449	Edwardsia
C 27	17.6	73.3	12372	1	AE013979	Yersinia
C 28	17.6	73.3	103647	8	CNS08CD1	Medicago
C 29	17.6	73.3	105229	8	AC149493	Medicago
C 30	17.6	73.3	110000	1	BX936398	Continuation (7 of
C 31	17.6	73.3	118625	8	AC122161	Medicago
C 32	17.6	73.3	125865	8	AC122171	Medicago
C 33	17.6	73.3	148899	2	BX511217	Danio rer
C 34	17.6	73.3	149008	2	BX296524	Danio rer
C 35	17.6	73.3	158035	2	AC149137	Xenopus t
C 36	17.6	73.3	214050	1	AJ414142	Yersinia
C 37	17.6	73.3	293371	1	AE017141	Yersinia
C 38	17.4	72.5	7309	6	CQ614000	Sequence
C 39	17.4	72.5	15792	2	AC014334	Drosophil
C 40	17.4	72.5	184272	3	AC009211	Drosophil
C 41	17.4	72.5	188967	3	AC007754	Drosophil
C 42	17.4	72.5	208424	3	AE003699	Drosophil
C 43	17.2	71.7	746	3	AY603562	Culicoides
C 44	17.2	71.7	10960	1	AE003349	Chlamydia
C 45	17.2	71.7	12345	1	AE006757	Sulfolobu

ALIGNMENTS

BD082356 24 bp DNA linear PAT 27-AUG-2002
76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
corresponding polynucleotide molecules.

ACCESSION BD082356.1 GI:22627966

VERSION JP 2001523954-A/63.

KEYWORDS Mastadenovirus

SOURCE Mastadenovirus

ORGANISM Viruses; dsDNA viruses, no RNA stage; Adenoviridae.

REFERENCE 1 (bases 1 to 24)

AUTHORS Kleanthous, H., Lissolo, L., Tomb, J.F., Miller, C. and Garawi, A.A.

TITLE 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
corresponding polynucleotide molecules

JOURNAL Patent: JP 2001523954-A 63 27-NOV-2001;

MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET

VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,

HUMAN GENOME SCIENCES INC

COMMENT PN JP 2001523954-A/63

PD 27-NOV-2001

PF 31-MAR-1998 JP 1998541962

PI HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI

MILLER,

PI AMAL AL GARAWI

PC A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,

PC A61K39/40,

PC G01N33/554, G01N33/569

CC Strandedness: Single;

CC Topology: Linear;

FH Key Location/Qualifiers.

1. 24

/organism="Mastadenovirus"

/mol_type="genomic DNA"

/db_xref="taxon:10509"

ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 24;

Best Local Similarity 100.0%; Pred. No. 2.1; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0;

```
QY 1 CATAACGCAAAATAACGCTACGCAT 24
|||||
Db 1 CATAACGCAAAATAACGCTACGCAT 24

RESULT 2
BD092341 1000 bp DNA linear PAT 27-AUG-2002
LOCUS Identification of polynucleotides encoding novel helicobacter
DEFINITION polyptides in the helicobacter genome.
ACCESSION BD092341
VERSION BD092341.1 GI:22637952
KEYWORDS JP 2001527393-A/122
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1000)
AUTHORS Kleanthous,H., Garawi,A.A., Miller,C., Tomb,J.F. and Oomen,R.P.
TITLE Identification of polynucleotides encoding novel helicobacter
JOURNAL polyptides in the helicobacter genome
Patent: JP 2001527393-A 122 25-DEC-2001;
MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET
VACCINS AGROBIOLOGICAL RESOURCES MINISTRY O SA, HUMAN GENOME
SCIENCES INC
COMMENT PN JP 2001527393-A/122
PD 25-DEC-2001
PF 01-APR-1998 JP 1998541947
PR 01-APR-1997 US 08/833457,24-JUN-1997 US 08/881227 PR
29-JUL-1997 US 08/902615
PI HAROLD KLEANTHOU,AMAL AL GARAWI,CHARLES MILLER,JEAN FRANCOIS
PI TOMB,
PI RAYMOND PETER OOMEN
PC A01N43/04,A61K31/70
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..1000
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 100.0%; Score 24; DB 6; Length 1000;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATAACGCAAAATAACGCTACGCAT 24
|||||
Db 123 CATAACGCAAAATAACGCTACGCAT 146

RESULT 3
AR342398 1149 bp DNA linear PAT 17-AUG-2003
LOCUS Sequence 23 from patent US 6576244.
DEFINITION AR342398
ACCESSION AR342398
VERSION AR342398.1 GI:33737371
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1149)
AUTHORS Weltzin,R.A. and Guy,B.
TITLE LT and CT in parental immunization methods against helicobacter
infection
JOURNAL Patent: US 6576244-A 23 10-JUN-2003;
FEATURES
source
1..1149
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
```

```
Query Match 100.0%; Score 24; DB 6; Length 1149;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATAACGCAAAATAACGCTACGCAT 24
|||||
Db 178 CATAACGCAAAATAACGCTACGCAT 201

RESULT 4
BD082346 1149 bp DNA linear PAT 27-AUG-2002
LOCUS 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
DEFINITION corresponding polynucleotide molecules.
ACCESSION BD082346
VERSION BD082346.1 GI:22627956
KEYWORDS JP 2001523954-A/53.
SOURCE Mastadenovirus
ORGANISM Mastadenovirus
REFERENCE 1 (bases 1 to 1149)
AUTHORS Kleanthous,H., Lissolo,L., Tomb,J.F., Miller,C. and Garawi,A.A.
TITLE 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
JOURNAL corresponding polynucleotide molecules
Patent: JP 2001523954-A 53 27-NOV-2001;
MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET
VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
HUMAN GENOME SCIENCES INC
COMMENT PN JP 2001523954-A/53
PD 27-NOV-2001
PF 31-MAR-1998 JP 1998541962
PI HAROLD KLEANTHOU,LING LISSOLO,JEAN FRANCOIS TOMB,CHARLES PI
MILLER,
PI AMAL AL GARAWI
PC A01N43/04,A01N59/16,A61K9/48,A61K31/70,A61K31/715,A61K39/02,
PC A61K39/40,
PC G01N33/554,G01N33/569
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..1149
/organism="Mastadenovirus"
/mol_type="genomic DNA"
/db_xref="taxon:10509"
ORIGIN
Query Match 100.0%; Score 24; DB 6; Length 1149;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATAACGCAAAATAACGCTACGCAT 24
|||||
Db 178 CATAACGCAAAATAACGCTACGCAT 201

RESULT 5
AE000538 11421 bp DNA linear BCT 06-APR-1999
LOCUS Helicobacter pylori 26695 section 16 of 134 of the complete genome.
DEFINITION AE000538 AE000511
ACCESSION AE000538.1 GI:2313263
VERSION AE000538.1
SOURCE Helicobacter pylori 26695
ORGANISM Helicobacter pylori 26695
REFERENCE 1 (bases 1 to 11421)
AUTHORS Tomb,J.-F., White,O., Kervilave,A.R., Clayton,R.A., Sutton,G.G.,
Fleischmann,R.D., Ketchum,K.A., Klein,H.P., Gill,S.,
Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D.,
```

Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
 Peterson,J.D., Kelley,J.M., Karp,P.D., Smith,H.O., Fraser,C.M. and
 Venter,J.C.

The complete genome sequence of the gastric pathogen *Helicobacter*

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

pylori
 Nature 388 (6642), 539-547 (1997)

97394467

9252185

2 (bases 1 to 11421)

Tomb,J.-F., White,O., Kerlavage,A.R., Clayton,R.A., Sutton,G.G.,
 Fleischmann,R.D., Ketchum,K.A., Klenk,H.P., Gill,S.,
 Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F.,
 Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khalak,H.G.,
 Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D.,
 Hickey,E.K., Berg,D.E., Gocayne,J.D., Utterback,T.R.,
 Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C.,
 Bowman,C., Watthey,L., Wallin,E., Hayes,W.S., Borodovsky,M.,
 Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.

Direct Submission

Submitted (06-AUG-1997) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

3 (bases 1 to 11421)

White.O.

Direct Submission

Submitted (17-MAR-1999) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. .11421

/organism="Helicobacter pylori 26695"

/mol_type="genomic DNA"

/strain="26695"

/db_xref="taxon:85962"

246. .1145

/gene="HP0175"

246. .1145

/gene="HP0175"

/note="similar to PID:671840 SP:Q46105 percent identity:
 34.90; identified by sequence similarity; putative"

/codon_start=1

/product="cell binding factor 2"

/protein_id="AAD07245.1"

/db_xref="GI:2313264"

/translation="MKKNILNLLVGLSTFLMAKPAHANNATHNTTKTDTSSAGV
 LATVDGRITKSDPMIKQPNFDFDLKEKEALIDQIRLALVNEAKTEKLDLS
 TPEKAMAEVKKOALVEFKAQAEVKKQIQPEKQMDFFVANKDOLFVKQEAHAR
 HILVTEAEKRLISEIDKQPKAKKEAFIELANDTIDPNSKNAQNGDGLGKQKQKQ
 MAPDFSKAALTPGDTYKTPVKTEFGYHIIYLISKOSFVITYTTEQAKPTINGMLQEK
 LFOERMNQRIBELRRKAKIVINK"

1162. .2085

/gene="HP0176"

1162. .2085

/gene="HP0176"

/note="similar to GB:M22039 SP:P13243 PID:460911
 PID:853765 GB:AL009126 percent identity: 45.96; identified
 by sequence similarity; putative"

/codon_start=1

/product="fructose-bisphosphate aldolase (tcr)"

/protein_id="AAD07246.1"

/db_xref="GI:2313265"

/translation="MLVKGNEILLKAKKEGYGVGFNFVPEMLNAIPEAGNEENSPIL
 FIOTSEGAIKYMGIDMAVMVKTWCERYPHIPVALHLDHGTTFESCEKAVKAGTFSVM
 IDASHAFRENLBELTSKVVMKHNAGVSVEALGELMGIEDNISVDEKDALVNPKEA
 EQFKVESQVDYLAPAGISGHAFFKPEKLODFERLQVEVKRLTNPILVHGAISADPN
 VRKSYLDKGGKSGVPFFLQESVKGKINKVNTDRLRIAFIAEVKRVKVAENKSDQ
 FDLRKFPSPAQLAKNVVVKRMKLGSAKNI"

2107. .2670

/gene="HP0177"

2107. .2670

/gene="HP0177"

/note="similar to GB:U14003 SP:P33398 GB:X61676 PID:433670
 PID:536991 percent identity: 45.05; identified by sequence

similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="translation elongation factor EF-P (efp)"
 /protein_id="AAD07247.1"
 /db_xref="GI:2313266"
 /translation="MAIGSELKGLKIELGGVPYRIVYQHVRFKGAAPVRAKIKS
 PLDGKVIKTEPHAGDKCEPNLVEKTKQYLYHDGDTYQFMDIESYQALINDSQVGEA
 SKMLDGMQVQLLNDKAIISVDVQVQVVAIVETAPNFKGTSSASKKPATLETGVA
 VQVFPFLVEGRIIKVNTETERYLEKVK"
 complement (3179. .4201)
 /gene="HP0178"
 complement (3179. .4201)
 /note="HP0178"
 /note="similar to GB:L77117 PID:1591717 percent identity:
 36.25; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="spore coat polysaccharide biosynthesis protein
 B"
 /protein_id="AAD07248.1"
 /db_xref="GI:2313267"
 /translation="WLOPPKIVAEISANHNODNLAKESLHAIKESGADPVKLOTYTP
 SCWTLNSKEDPPIIQGLTMDKENLYELQKASTQKPIILSSGATHTLQDAISLCR
 SQALLESNCPMKIKASPEIVOLDLIEKARTQKPIILSSGATHTLQDAISLCR
 RVNFDITLLKCVSAIPSKIBDANLLSMVLGEIFGVKFGLSHDITIGSLCPILATTLG
 ASMIKHFILNLSIQTPDSAFNGFNGKSMVEAIKQSVLALGEEPRINPKTLKRR
 FFARSLFVINDIQKGEALTENNIALRPNLGLHPKFKYKEILGQKASKFLKANTPLSAD
 DIERSL"
 complement (4205. .4846)
 /gene="HP0179"
 complement (4205. .4846)
 /note="HP0179"
 /note="similar to GP:1786703 percent identity: 37.62;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="ABC transporter, ATP-binding protein"
 /protein_id="AAD07249.1"
 /db_xref="GI:2313268"
 /translation="MIKAINISHAPEKPLVNGVNLHKPKESLAILGVSGSKSTLLS
 HLAITMKNPSTISLEHDIYALNSKKLELRRLKVGIIIFQSHYLFKGPBALENLQV
 ASILAKQENHSLLEQLGIAHTLQGVGELSGQQQRLSIARLSKPKRIIADSEPTG
 NLDITTSANVISMLQNTYIETKEGALVLATHEHLAFTCSQVYRLEKEVLEIKK"
 complement (4843. .6120)
 /gene="HP0180"
 complement (4843. .6120)
 /note="similar to GB:L42023 SP:P44626 PID:1003497
 PID:1222229 PID:1204555 percent identity: 28.03;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="apolipoprotein N-acyltransferase (cute)"
 /protein_id="AAD07250.1"
 /db_xref="GI:2313269"
 /translation="WRLLFNQNAFLACMEVSSVYVNAVLDVAIENPYISITLTSL
 LAPLSMLAFLKTPNSAFALGFVGCALLFYWCALSFRYSDFTYLLPLIIVLIALVGV
 LFYLLLYFENPYFRLLSFLGSSFIHPFGPDMVDPDSFYSVFRVDKLSLSLFLACI
 FLSTPKLKKRIIIGVLLGLDNGFKTSLDKKVGNIELVSTKTPQDLKEDSSYLDQ
 TENNILEIKLQAQKQTLIVFPETAYPIALENSPFKAKLEDSDNIAILIGTLRTQG
 YNLNSSFLESKESQVADKVLAPFGTETMPLFPLQKLEKLPFGESTYLYRNAPHF
 SDFTLDFTPLPLICYGTSKPAYSNSPKIFIVMSNNANFSPSIEPTLQRTLLKYVA
 RYVDKIILHSANFSTSYILSPSLGLDILFRKRS"
 6378. .7052
 /gene="HP0181"
 6378. .7052
 /note="HP0181"
 /note="hypothetical protein; identified by GeneMark;
 putative"
 /codon_start=1
 /transl_table=11
 /product="H. pylori predicted coding region HP0181"

```
/protein_id="AAD07253.1"
/db_xref="GI:2313272"
/translation="MVVAFGIRGFYHGFSVSEVAGTIGLIVGVYASRYSVAVGNLFS
EHLVDLRNETWNLGFLGLVLASIWVFLAFGLGKVLVFGSLGIDKALGFIPLSCL
KTFVLVSLFIALSKMEYMKDANAYLOKSAFFFTKMSVASKIMRLDGVKRVQNLKD
NLEMSDSDVKNKESFNKAKMDKGVSEKKEKADLPKNMLDPKANQTPNPPTP
SNKEPL"
```

gene

7062..8567

CDS

/gene="HP0182"

7062..8567

/gene="HP0182"

```
/note="similar to GB:M63448 SP:P41258 PID:144208 percent
identity: 58.63; identified by sequence similarity;
putative"
```

/codon_start=1

/transl_table=11

/product="lysyl-tRNA synthetase (lyss)"

/protein_id="AAD07251.1"

/db_xref="GI:2313270"

```
/translation="MFSNQYIQORIHKANSIREGKNPYQNGLKRSLTNAAPLEKYAV
VKGLEPKDKKCSIVGRVKLLRMGMKACFIKVEDESTILOVYVQNEINDEPKSLK
KHLEVGDIVLVKGFPAFTKGELSIHALEFHILSKTIIVLPKEKFGHSDIELRQRY
LDLVNPSVDVFKKRSILVSVRKFEMEGFLEVETPMHPPIFGANARPFITYHNA
LEVERYLRIAPELYLKRLIVGFAVFEINERFNEGMDHSHNPEFTMIEFYWHTY
EDLIELSKRLPDYLLKTLNLSKIIYNDMEVDFNQTSVISYLDALETIGGISKDILEK
EDRLAYLLEOGIKVEPNLTGKLLAEAFDHFVEHOLINPTFVQYPLIEISPLARRND
SNPNIADFEPIAGKEANGFSELNDPLDLQERFKNQVAEKEGEDEAQQMDDEYVW
ALAHGMPTAGQIGIDRLVMLLTGAKSIKDVILFPAMPVPKNDPNVSESE"
```

gene

8567..9817

CDS

/gene="HP0183"

8567..9817

/gene="HP0183"

```
/note="similar to GB:J01620 SP:P00477 GB:V00283 PID:146218
PID:41603 percent identity: 53.96; identified by sequence
similarity; putative"
```

/codon_start=1

/transl_table=11

/product="serine hydroxymethyltransferase (glyA)"

/protein_id="AAD07252.1"

Query Match

Best Local Similarity 100.0%; Score 24; DB 1; Length 11421;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 CATAACGCAATAACGCTACGCAT 24

Db

318 CATAACGCAATAACGCTACGCAT 341

RESULT 6

BD061716

LOCUS

BD061716 1082 bp DNA linear PAT 27-AUG-2002

DEFINITION

Antigenic composition and method of detection for Helicobacter

Pylori.

ACCESSION

BD061716

VERSION

BD061716.1 GI:22607321

KEYWORDS

JP 2001517091-A/50.

SOURCE

synthetic construct

ORGANISM

other sequences; artificial sequences.

REFERENCE

1 (bases 1 to 1082)

AUTHORS

Chow,T.P., Fry,K.E., Lim,M.Y. and Meates,C.P.

TITLE

Antigenic composition and method of detection for Helicobacter

JOURNAL

Patent: JP 2001517091-A 50 02-OCT-2001;

COMMENT

PN JP 2001517091-A/50

FEATURES

source

Location/Qualifiers

1..1082

/organism="synthetic construct"

/mol_type="genomic DNA"

/db_xref="taxon:32630"

ORIGIN

Query Match 93.3%; Score 22.4; DB 6; Length 1082;

Best Local Similarity 95.8%; Pred. No. 8.5;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CATAACGCAATAACGCTACGCAT 24

Db 237 CATAACGCAATAACGCTACGCAT 260

RESULT 7

BD061969

LOCUS

BD061969 1082 bp DNA linear PAT 27-AUG-2002

DEFINITION

Antigenic composition and method of detection for Helicobacter

Pylori.

ACCESSION

BD061969.1 GI:22607574

VERSION

JP 2001517091-A/303.

KEYWORDS

synthetic construct

SOURCE

synthetic construct

ORGANISM

other sequences; artificial sequences.

REFERENCE

1 (bases 1 to 1082)

AUTHORS

Chow,T.P., Fry,K.E., Lim,M.Y. and Meates,C.P.

TITLE

Antigenic composition and method of detection for Helicobacter

JOURNAL

Patent: JP 2001517091-A 303 02-OCT-2001;

COMMENT

PN JP 2001517091-A/303

PD 02-OCT-2001

PF 25-APR-1998 JP 1998547263

PR 25-APR-1997 US 60/045107,14-OCT-1997 US 60/061958 PI

THERESA P CHOW,KIRK E FRY,MOON Y LIM,C P MCATEE PC

CI2N15/31,C07K14/205,C07K16/12,A61K39/106

CC Strandedness: Single;

CC Topology: Linear;

CC Topology: Linear;

FEATURES

source

Location/Qualifiers

1..1082

/organism="synthetic construct"

/mol_type="genomic DNA"

/db_xref="taxon:32630"

ORIGIN

Query Match 93.3%; Score 22.4; DB 6; Length 1082;

Best Local Similarity 95.8%; Pred. No. 8.5;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CATAACGCAATAACGCTACGCAT 24

Db 237 CATAACGCAATAACGCTACGCAT 260

RESULT 8

BD061705

LOCUS

BD061705 1117 bp DNA linear PAT 27-AUG-2002

DEFINITION

Antigenic composition and method of detection for Helicobacter

Pylori.

ACCESSION

BD061705.1 GI:22607310

VERSION

JP 2001517091-A/39.

KEYWORDS

synthetic construct

SOURCE

synthetic construct

ORGANISM

other sequences; artificial sequences.

REFERENCE

1 (bases 1 to 1117)

AUTHORS

Chow,T.P., Fry,K.E., Lim,M.Y. and Meates,C.P.

TITLE

Antigenic composition and method of detection for Helicobacter

JOURNAL

Patent: JP 2001517091-A 39 02-OCT-2001;


```

GENELABS TECHNOLOGIES INC
PN JP 2001517091-A/39
PD 02-OCT-2001
PF 25-APR-1997 US 60/045107,14-OCT-1997 US 60/061958 PI
PR 25-APR-1997 US 60/045107,14-OCT-1997 US 60/061958 PI
THERESA P CHOW, KIRK E FRY, MOON Y LIM, C P MCATEE PC
C12N15/31, C07K14/205, C07K16/12, A61K39/106
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
    source
        1..1117
        /organism="synthetic construct"
        /mol_type="genomic DNA"
        /db_xref="taxon:32630"

ORIGIN
    Query Match 93.3%; Score 22.4; DB 6; Length 1117;
    Best Local Similarity 95.8%; Pred. No. 8.4;
    Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CATAACGCAATACGCTACGCAT 24
    |||||
Db 255 CATAACGCAATACGCTACGCAT 278

RESULT 9
LOCUS AE001454
DEFINITION Helicobacter pylori, strain J99 section 15 of 132 of the complete genome.
ACCESSION AE001454
VERSION AE001454.1
KEYWORDS GI:4154666
SOURCE Helicobacter pylori J99
ORGANISM Helicobacter pylori J99
REFERENCE 1 (bases 1 to 10085)
AUTHORS Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F. and Trust, T.J.
TITLE Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
JOURNAL Nature 397 (6715), 176-180 (1999)
MEDLINE 99120557
PUBMED 9923682
REFERENCE 2 (bases 1 to 10085)
AUTHORS King, B.L., Alm, R.A. and Trust, T.J.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA
COMMENT Addresses all correspondence to: hp@arch.us.astra.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ARCB's World Wide Web site. (URL: http://www.astra-boston.com/hpylori).

FEATURES
    source
        1..10085
        /organism="Helicobacter pylori J99"
        /mol_type="genomic DNA"
        /strain="J99"

```

```

/db_xref="taxon:85963"
complement (235..708)
/gene="jhp0153"
complement (235..708)
/gene="jhp0153"
note="similar to H. pylori 26695 gene HP0167"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05736.1"
/db_xref="GI:4154669"
/transl_table="MAGMELKNIISSETLNEIEKMAKTIIDGFPNCTAQTPTSPKTPSH
LQNTPSKNANTPLEPNAKIEQKEITEKEAKEIITEIIVQENPTQALINERV
FLKGLERTLVFKGMQALKEEALKRLDLVARFLQYQLSVLEKRELSERENTE"
complement (687..950)
/gene="jhp0154"
complement (687..950)
/gene="jhp0154"
note="similar to H. pylori 26695 gene HP0168"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05737.1"
/db_xref="GI:4154670"
/transl_table="MPLTITLAIHYEEQGFEEALQIYNILKKTDPDNTALQMKR
LEKIQKNCAPFKHDALLKQHYLNFIKGRLSVENLEKWLVEWN"
complement (997..2265)
/gene="jhp0155"
complement (997..2265)
/gene="jhp0155"
note="similar to H. pylori 26695 gene HP0169"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05738.1"
/db_xref="GI:4154671"
/transl_table="MNQVELSPAGNKKULKIALNYGADAVGVGVSHFSLNRACKEP
TLETFEGIDYAHALDKKYVATNGFPNQLKLEBHLKMALEPDAFIIAAGPVI
KLASKIAPHIPHLSTQANTVLTDAQVFDLGVKRVICARELSNDAVEIKKALPD
ELETFVHGSMCFAPSGCLISALQGVNPGNSCANDCRFDYEVYKNDPVMRLV
EEEGVGTIFNAKDLNLSGHIAETILSSNAISALAKIEGRTKSSYAAOTTRLYRLAVD
FYHNTLSPYSASELNTLKNRGFTDGYLMRPFERLDTQHOTAISGDFQVNGEITE
DGRFACKFTTTTAYEITAPKNAIITPVNDIGKIYTFEKRSYLVLYKILENNT
LETTHSGNVLRLPAPLPAFLPFLRTQVRV"
complement (2268..2999)
/gene="jhp0156"
complement (2268..2999)
/gene="jhp0156"
note="similar to H. pylori 26695 gene HP0170"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05739.1"
/db_xref="GI:4154672"
/transl_table="MTQEBDALMSGDLNLEALEAKEAKEKEPTKEPKESKNSNY
SERKTVKEDAEKYGKISPEHPPTTEHKVHQLDDVTRDEVKATQIFDQLDLI
GAASAEKAKVKTQEPHQKEIFDNLHAFPHVESFKTKALNEQQETLNAKSTEE
AANCSDSMQAMDIMQFDTHROKIERVAVNVMRLSQVMNSLPEGKIDDSKRVSSATY
ITGDDDKDLASADDIEALISFGAK"
complement (3069..4160)
/gene="prfB"
note="synonym: jhp0157"
complement (3069..4160)
/gene="prfB"
note="similar to H. pylori 26695 gene HP0171"
/codon_start=1
/transl_table=11
/product="PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2)"
/protein_id="AAD05740.1"
/db_xref="GI:4154673"
/transl_table="MDNYTYSLELKSQNKCDNIALIKPEKIKQELERIEKEQEDPN
FWQDLVKARDTNKEKRLLETYQKMSLSDSEVELFELAQNDSEVTLUSLYEEA
PTLHESVQKVEIEMLSGENDASNAIITQFAGGAGTESQDWSILYRMYLWAEKRSF

```

KSEILDYQGBEAGIKGVAFIIGENAYGLKNGVHRLVRIISPFDANAKRHTSFAS
VOISPELDDDIIEIDEXDQVYDVYRASSAGSQHVNTKTESAVRIITHPTGIIVQOQND
RSQHNKASALMKLSKYLELELEKQSSAKNEEKSEIGWGHQIRSYVLYPQYKQDA
RSHATSNVEAILDDGDIIDALEGLVLIARA"
/gene="moea"
/note="synonym: jhp0158"
complement (4221..5393)
/gene="moea"
complement (4221..5393)
/note="similar to H. pylori 26695 gene HP0172"
/codon_start=1
/transl_table=11
/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
/protein_id="AAD05741.1"
/db_xref="GI:4154674"

/translation="MISPEALKIHSNIPLKPLEVEVVSLEFESAGRILAEIICVHAL
PKFNQASMDGFGKMDIQKQTVIHIQIHPAGDQVSALEKNECVKIMTGAMVPGKIE
TIVTECMLESHGKDFALAPKDFKIHANIRQKGENASLNVLPKNTRLNYGHIALIAS
QGKREIKAPRKLIALFSSGDELPLGQNALECOVYDVNSGVFNMLKNYNTHFLGVL
KDDNLQIKILELOQYDVLASAGVSGDKDFKDALKERNALFYVEKVNLPKPKPVT
LAQLNQSLIIGLPNPLSCLVLRVLIIPLLERLSLNDKDFKLKPKFQAQINAPLKLNNK
RTHLLIGNSHQFIPIPNRNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFPEN"
complement (5403..6170)
/gene="flir"
/note="synonym: jhp0159"
complement (5403..6170)
/gene="flir"
/note="similar to H. pylori 26695 gene HP0173"
/codon_start=1
/transl_table=11
/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
/protein_id="AAD05742.1"
/db_xref="GI:4154675"

/translation="MLDPIQLSTPHVRDFFLLFURVSGVLSFPPFFPHLVLPSVRG
ALSIVAFIYPTFEISNAIYTPGFIACELFLGVCASVFIQIVFASIVFATDSI
SFSMLTASAYDSQKPIVQALILLAILLIDLSFHHQIILFVHDSUKAVPLG
RVFPEPEKALIVAFSHLFIQVSGMAFPICLVLLSDIIFGMIMKTHPQNLLAIGF
PVKAIGFVGIIILIASATMGREKEISLAFSVISKIF"
complement (6164..6892)
/gene="jhp0160"
complement (6164..6892)
/gene="jhp0160"
/note="similar to H. pylori 26695 gene HP0174"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05743.1"
/db_xref="GI:4154676"

/translation="MVLFLSIKKSFNDFLGRMLLNLGPIILLSLAPFGAIFYVNGG
SIVNYCOTLLQSLNDYASQGFAGVFAWFKALVYFLIFWIVLILLSVINIASIF
YTPLVSVLHOKYYPHVVLBERGSLFSIKYFLKSLTFLMLFLAVLTPLYFIPIGVF
GVFISVPHFLUFFKNTMSLDIASMIFNHQSYQNLKQHLKHYRFPFCYILFSLIPFF
NFFATLLQTLMTHYFFIFKKEC"
7361..8260
/gene="jhp0161"
7361..8260
/gene="jhp0161"
/note="similar to H. pylori 26695 gene HP0175"
/codon_start=1
/transl_table=11
/product="putative PRPTIDYL-PROLYL CIS-TRANS ISOMERASE"
/protein_id="AAD05744.1"
/db_xref="GI:4154677"

/translation="MKNILNLALVGLSASFLMAKPAHNANNSTHNTKTTDASAGV
LATVDGRIITSDSDMIKORPNFDFKLKEKEALIEQAIRTALVNEAKELNQ
TPEKAMAEQKALVEFWAKQAEVKKIQIPEKEMQDFYNAKMDQLFVKQFAHAR
HILVKTEBAKRIISEIDQPKAKKEAKFIELANDTIDPNSKNAQNGDGLKFKQKQ
MAPDESKAFAITPGDYTKTPUKTEFGYHIIYLISKDSPTYTYEQAKPTIKGLMKEK
LFQERNORISELRKHAKIVINK"
8274..9197
/gene="fba"
/note="synonym: jhp0162"
8274..9197

/gene="fba"
/note="similar to H. pylori 26695 gene HP0176"
/codon_start=1
/transl_table=11
/product="FRUCTOSE-BISPHOSPHATE ALDOLASE"
/protein_id="AAD05734.1"
/db_xref="GI:4154667"
/translation="MLVKGNEILLKAHKEGVGAFNFVNFEMLNALFEAGNEENSP
FIQASEGAIKYMGIDMAVGVMKINCERYPHIPVALHLDHGTTFPSCSEKAVKAGTSM

Query Match 80.0%; Score 19.2; DB 1; Length 10085;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CATAACGCCAATAACGCTACGCAT 24
||||| ||||| ||||| ||||| |||||
Db 7433 CATAATCGGAATACTCTACGCAT 7456

RESULT 10
AE013426
LOCUS Methanosarcina mazei strain Goel, section 208 of 379 of the
DEFINITION complete genome.
ACCESSION AE013426 AE008384
VERSION AE013426.1 GI:20906421
KEYWORDS Methanosarcina mazei Goel
SOURCE Methanosarcina mazei Goel
ORGANISM Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
Methanosarcinaceae; Methanosarcina.

REFERENCE 1 (bases 1 to 11193)
AUTHORS Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
Martinez-Arias,R., Henne,A., Wierer,A., Baumer,S., Jacobi,C.,
Brueggemann,H., Lienard,T., Christmann,A., Boemcke,M., Steckel,S.,
Bhattacharya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
The genome of Methanosarcina mazei: evidence for lateral gene
transfer between bacteria and archaea
J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002)

2 (bases 1 to 11193)
AUTHORS Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,
Martinez-Arias,R., Henne,A., Wierer,A., Baumer,S., Jacobi,C.,
Brueggemann,H., Lienard,T., Christmann,A., Boemcke,M., Steckel,S.,
Bhattacharya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,
Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.
Direct Submission
Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute
for Microbiology and Genetics, Grisebachstrasse 8, Goettingen
37077, Germany

FEATURES
source
1..11193
/organism="Methanosarcina mazei Goel"
/mol_type="genomic DNA"
/strain="Goel"
/db_xref="taxon:192952"
complement (131..1669)
/gene="MM1892"
complement (131..1669)
/gene="MM1892"
/EC_number="3.1.6.1"
/codon_start=1
/transl_table=11
/product="Arylsulfatase"
/protein_id="AAM31588.1"
/db_xref="GI:20906422"

FEATURES
source
1..11193
/organism="Methanosarcina mazei Goel"
/mol_type="genomic DNA"
/strain="Goel"
/db_xref="taxon:192952"
complement (131..1669)
/gene="MM1892"
complement (131..1669)
/gene="MM1892"
/EC_number="3.1.6.1"
/codon_start=1
/transl_table=11
/product="Arylsulfatase"
/protein_id="AAM31588.1"
/db_xref="GI:20906422"

FEATURES
source
1..11193
/organism="Methanosarcina mazei Goel"
/mol_type="genomic DNA"
/strain="Goel"
/db_xref="taxon:192952"
complement (131..1669)
/gene="MM1892"
complement (131..1669)
/gene="MM1892"
/EC_number="3.1.6.1"
/codon_start=1
/transl_table=11
/product="Arylsulfatase"
/protein_id="AAM31588.1"
/db_xref="GI:20906422"

FEATURES
source
1..11193
/organism="Methanosarcina mazei Goel"
/mol_type="genomic DNA"
/strain="Goel"
/db_xref="taxon:192952"
complement (131..1669)
/gene="MM1892"
complement (131..1669)
/gene="MM1892"
/EC_number="3.1.6.1"
/codon_start=1
/transl_table=11
/product="Arylsulfatase"
/protein_id="AAM31588.1"
/db_xref="GI:20906422"

FEATURES
source
1..11193
/organism="Methanosarcina mazei Goel"
/mol_type="genomic DNA"
/strain="Goel"
/db_xref="taxon:192952"
complement (131..1669)
/gene="MM1892"
complement (131..1669)
/gene="MM1892"
/EC_number="3.1.6.1"
/codon_start=1
/transl_table=11
/product="Arylsulfatase"
/protein_id="AAM31588.1"
/db_xref="GI:20906422"


```

Db      31369 ATACGAAAATACGATAGCAT 31391
||||| ||||| ||||| || |||||
LOCUS   AC010922      152607 bp      DNA      linear      INV 31-JUL-2004
DEFINITION Drosophila melanogaster clone BACR09H01, complete sequence.
ACCESSION AC010922
VERSION   AC010922.10 GI:50872288
KEYWORDS HTG.
SOURCE    Drosophila melanogaster (fruit fly)
ORGANISM  Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
          1 (bases 1 to 152607)
REFERENCE Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
AUTHORS   Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
          Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
          Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
          Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
          Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
          Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
          Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
          Rubin,G.M.
          Direct Submission
          Submitted (28-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
          Laboratory, MS 64-121, Berkeley, CA 94720, USA
          2 (bases 1 to 152607)
REFERENCE Celniker,S., Carlson,J., Wan,K., Pfeiffer,B., Frise,E., George,R.,
AUTHORS   Hoskins,R., Stapleton,M., Pacleb,J., Park,S., Svirskas,R.,
          Smith,E., Yu,C. and Rubin,G.
          Direct Submission
          Submitted (31-JUL-2004) Berkeley Drosophila Genome Project, MS
          64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
          Berkeley, CA 94720, US
          On Jul 31, 2004 this sequence version replaced gi:13957598.
          Sequence submitted by:
          Berkeley Drosophila Genome Project
          Lawrence Berkeley National Laboratory
          Berkeley, CA 94720
          This sequence submission incorporates changes made during
          reevaluation of the assembly or fingerprint verification of the
          clone. For further information about this sequence, including its
          location and relationship to other sequences, please visit our
          sequence archive Web site (http://www.fruitfly.org/sequence/) or
          send email to bdgp@fruitfly.org.
          Location/Qualifiers
          FEATURES
            source
              1..152607
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /strain="y; cn bw sp"
                /db_xref="taxon:7227"
                /chromosome="X"
                /map="15B-15B"
                /clone="BAC clone BACR09H01 (D896)"
                /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                Drosophila melanogaster BAC library, partial ECORI in
                pBACE3.6)"
          ORIGIN
            Query Match      75.8%; Score 18.2; DB 3; Length 152607;
            Best Local Similarity 87.0%; Pred. No. 5.3e+02;
            Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

            Qy      2 ATACGAAAATACGCTACGCAT 24
            ||||| ||||| ||||| |||||
            Db      49945 ATACGAAAATACGATAGCAT 49923

            RESULT 12
            LOCUS   AC010922/c      152607 bp      DNA      linear      INV 31-JUL-2004
            DEFINITION Drosophila melanogaster clone BACR09H01, complete sequence.
            ACCESSION AC010922/c
            VERSION   AC010922.10 GI:50872288
            KEYWORDS HTG.
            SOURCE    Drosophila melanogaster (fruit fly)
            ORGANISM  Drosophila melanogaster
                    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                    Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                    Ephydroidea; Drosophilidae; Drosophila.
                    1 (bases 1 to 152607)
            REFERENCE Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            AUTHORS   Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
                    Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
                    Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
                    Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
                    Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
                    Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
                    Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
                    Rubin,G.M.
                    Direct Submission
                    Submitted (28-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
                    Laboratory, MS 64-121, Berkeley, CA 94720, USA
                    2 (bases 1 to 152607)
            REFERENCE Celniker,S., Carlson,J., Wan,K., Pfeiffer,B., Frise,E., George,R.,
            AUTHORS   Hoskins,R., Stapleton,M., Pacleb,J., Park,S., Svirskas,R.,
                    Smith,E., Yu,C. and Rubin,G.
                    Direct Submission
                    Submitted (31-JUL-2004) Berkeley Drosophila Genome Project, MS
                    64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
                    Berkeley, CA 94720, US
                    On Jul 31, 2004 this sequence version replaced gi:13957598.
                    Sequence submitted by:
                    Berkeley Drosophila Genome Project
                    Lawrence Berkeley National Laboratory
                    Berkeley, CA 94720
                    This sequence submission incorporates changes made during
                    reevaluation of the assembly or fingerprint verification of the
                    clone. For further information about this sequence, including its
                    location and relationship to other sequences, please visit our
                    sequence archive Web site (http://www.fruitfly.org/sequence/) or
                    send email to bdgp@fruitfly.org.
                    Location/Qualifiers
                    FEATURES
                      source
                        1..152607
                          /organism="Drosophila melanogaster"
                          /mol_type="genomic DNA"
                          /strain="y; cn bw sp"
                          /db_xref="taxon:7227"
                          /chromosome="X"
                          /map="15B-15B"
                          /clone="BAC clone BACR09H01 (D896)"
                          /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                          Drosophila melanogaster BAC library, partial ECORI in
                          pBACE3.6)"
                    ORIGIN
                      Query Match      75.8%; Score 18.2; DB 3; Length 152607;
                      Best Local Similarity 87.0%; Pred. No. 5.3e+02;
                      Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

                      Qy      2 ATACGAAAATACGCTACGCAT 24
                      ||||| ||||| ||||| |||||
                      Db      49945 ATACGAAAATACGATAGCAT 49923

                      RESULT 13
                      LOCUS   AC010921/c      298641 bp      DNA      linear      INV 15-MAR-2004
                      DEFINITION Drosophila melanogaster chromosome X, section 56 of 74 of the
                      complete sequence.
                      ACCESSION AE003504
                      VERSION   AE003504.3 GI:22832424
                      KEYWORDS

```

REFERENCE	4 (bases 1 to 298641)	REFERENCE	4 (bases 1 to 298641)
AUTHORS	Kaminker,J.S., Bergman,C.M., Krommiller,B., Carlson,J., Svirskaas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E., Rubin,G.M., Ashburner,M. and Celniker,S.E.	AUTHORS	Kaminker,J.S., Bergman,C.M., Krommiller,B., Carlson,J., Svirskaas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E., Rubin,G.M., Ashburner,M. and Celniker,S.E.
TITLE	The transposable elements of the <i>Drosophila melanogaster</i> euchromatin: a genomics perspective	TITLE	The transposable elements of the <i>Drosophila melanogaster</i> euchromatin: a genomics perspective
JOURNAL	Genome Biol. 3 (12), RESEARCH0084 (2002)	JOURNAL	Genome Biol. 3 (12), RESEARCH0084 (2002)
MEDLINE	22426070	MEDLINE	22426070
PUBMED	12537573	PUBMED	12537573
REFERENCE	5 (bases 1 to 298641)	REFERENCE	5 (bases 1 to 298641)
AUTHORS	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	AUTHORS	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE	Direct Submission	TITLE	Direct Submission
JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA	JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
REFERENCE	6 (bases 1 to 298641)	REFERENCE	6 (bases 1 to 298641)
AUTHORS		AUTHORS	
CONSTRM	FlyBase	CONSTRM	FlyBase
TITLE	Direct Submission	TITLE	Direct Submission
JOURNAL	Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA	JOURNAL	Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA
REFERENCE	7 (bases 1 to 298641)	REFERENCE	7 (bases 1 to 298641)
AUTHORS		AUTHORS	
CONSTRM	FlyBase	CONSTRM	FlyBase
TITLE	Direct Submission	TITLE	Direct Submission
JOURNAL	Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA	JOURNAL	Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA
COMMENT	On Sep 13, 2002 this sequence version replaced gi:10728299.	COMMENT	On Sep 13, 2002 this sequence version replaced gi:10728299.
FEATURES	Location/Qualifiers	FEATURES	Location/Qualifiers
source	1..298641	source	1..298641
gene	/organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /chromosome="X" complement(651..2980) /gene="wus" /locus_tag="CG9089" /map="15C4-15C4" /db_xref="FLYBASE:FBgn0030805" complement(join(651..2145,2658..2980)) /gene="wus" /locus_tag="CG9089" /product="CG9089-RA" /db_xref="FLYBASE:FBgn0030805" complement(907..2127) /gene="wus" /locus_tag="CG9089" /codon_start=1 /product="CG9089-PA" /protein_id="AAF48683.1" /db_xref="GI:7293303" /db_xref="FLYBASE:FBgn0030805" /translation="MGTQTVSPVAGKDKTNDTSKGSPSKAKSNGAOKTKSKWNA LPQKSVVYALCWLFGIFGLHHLYLHRDRHAFIMWTLGGYNGICGWGELFLIPEY VRDANEOPKVFYKMFVAKLQPKPPYSPKQVGMIGHLFGYCSMAIPETLVGGMD LSFLHWTIPLFVSLGVNLGNIGREQGWWHCLVAAYLARYLYDITYSLTLLG VAALTFDGLSKQWRRTTPRGSGPGRFKLTITAVIYICAFWGSFLYFNGTISDDEGE VPIEATHNFIPLASAWTDLQALLDTVYQAHHGWETWKEVFSMDVDCGRNSYKVL GVSATASQAETITAYARKLSKEYHPDKVDEGLRAQHQRFIEIQQAYSVLISKSKNSR RKNQYQEEEAIVL"	gene	/organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /chromosome="X" complement(651..2980) /gene="wus" /locus_tag="CG9089" /map="15C4-15C4" /db_xref="FLYBASE:FBgn0030805" complement(join(651..2145,2658..2980)) /gene="wus" /locus_tag="CG9089" /product="CG9089-RA" /db_xref="FLYBASE:FBgn0030805" complement(907..2127) /gene="wus" /locus_tag="CG9089" /codon_start=1 /product="CG9089-PA" /protein_id="AAF48683.1" /db_xref="GI:7293303" /db_xref="FLYBASE:FBgn0030805" /translation="MGTQTVSPVAGKDKTNDTSKGSPSKAKSNGAOKTKSKWNA LPQKSVVYALCWLFGIFGLHHLYLHRDRHAFIMWTLGGYNGICGWGELFLIPEY VRDANEOPKVFYKMFVAKLQPKPPYSPKQVGMIGHLFGYCSMAIPETLVGGMD LSFLHWTIPLFVSLGVNLGNIGREQGWWHCLVAAYLARYLYDITYSLTLLG VAALTFDGLSKQWRRTTPRGSGPGRFKLTITAVIYICAFWGSFLYFNGTISDDEGE VPIEATHNFIPLASAWTDLQALLDTVYQAHHGWETWKEVFSMDVDCGRNSYKVL GVSATASQAETITAYARKLSKEYHPDKVDEGLRAQHQRFIEIQQAYSVLISKSKNSR RKNQYQEEEAIVL"
mRNA		mRNA	
CDS		CDS	
gene	complement(3234..5016) /locus_tag="CG13001" /map="15C4-15C4" /db_xref="FLYBASE:FBgn0030806" complement(3234..5016) /locus_tag="CG13001" /product="CG13001-RA" /db_xref="FLYBASE:FBgn0030806" complement(3729..4736) /locus_tag="CG13001" /codon_start=1 /product="CG13001-PA" /protein_id="AAF48684.1" /db_xref="GI:7293304"	gene	complement(3234..5016) /locus_tag="CG13001" /map="15C4-15C4" /db_xref="FLYBASE:FBgn0030806" complement(3234..5016) /locus_tag="CG13001" /product="CG13001-RA" /db_xref="FLYBASE:FBgn0030806" complement(3729..4736) /locus_tag="CG13001" /codon_start=1 /product="CG13001-PA" /protein_id="AAF48684.1" /db_xref="GI:7293304"
mRNA		mRNA	
CDS		CDS	
TITLE	Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a systematic review	TITLE	Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a systematic review
JOURNAL	Genome Biol. 3 (12), RESEARCH0083 (2002)	JOURNAL	Genome Biol. 3 (12), RESEARCH0083 (2002)
MEDLINE	22426069	MEDLINE	22426069
PUBMED	12537572	PUBMED	12537572
REFERENCE	3 (bases 1 to 298641)	REFERENCE	3 (bases 1 to 298641)
AUTHORS	Misra,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S., Hradecky,P., Huang,Y., Kaminker,J.S., Millburn,G.H., Prochnik,S.E., Smith,C.D., Tupy,J.L., Whitfield,E.J., Bayraktaroglu,L., Berman,B.P., Bettencourt,B.R., Celniker,S.E., de Grey,A.D., Drysdale,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J., Shu,S.O., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,W.M., Rubin,G.M. and Lewis,S.E.	AUTHORS	Misra,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S., Hradecky,P., Huang,Y., Kaminker,J.S., Millburn,G.H., Prochnik,S.E., Smith,C.D., Tupy,J.L., Whitfield,E.J., Bayraktaroglu,L., Berman,B.P., Bettencourt,B.R., Celniker,S.E., de Grey,A.D., Drysdale,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J., Shu,S.O., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,W.M., Rubin,G.M. and Lewis,S.E.
TITLE	Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a systematic review	TITLE	Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a systematic review
JOURNAL	Genome Biol. 3 (12), RESEARCH0079 (2002)	JOURNAL	Genome Biol. 3 (12), RESEARCH0079 (2002)
MEDLINE	22426065	MEDLINE	22426065
PUBMED	12537568	PUBMED	12537568
REFERENCE	2 (bases 1 to 298641)	REFERENCE	2 (bases 1 to 298641)
AUTHORS	Celniker,S.E., Wheeler,D.A., Krommiller,B., Carlson,J.W., Halpern,A., Patel,S., Adams,M., Champe,M., Dugan,S.P., Frise,E., Hodgson,A., George,R.A., Hoskins,R.A., Laverty,T., Muzny,D.M., Nelson,C.R., Pacleib,J.M., Park,S., Pfeiffer,B.D., Richards,S., Sotengen,E.J., Svirskaas,R., Tabor,P.E., Wan,K., Stapleton,M., Sutton,G.G., Venter,C., Weinstein,G., Scherer,S.E., Myers,E.W., Gibbs,R.A. and Rubin,G.M.	AUTHORS	Celniker,S.E., Wheeler,D.A., Krommiller,B., Carlson,J.W., Halpern,A., Patel,S., Adams,M., Champe,M., Dugan,S.P., Frise,E., Hodgson,A., George,R.A., Hoskins,R.A., Laverty,T., Muzny,D.M., Nelson,C.R., Pacleib,J.M., Park,S., Pfeiffer,B.D., Richards,S., Sotengen,E.J., Sv

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 17:27:10 ; Search time 804.69 Seconds
(without alignments)
1866.696 Million cell updates/sec

Title: US-10-039-183A-15

Perfect score: 31

Sequence: 1 gctaattatttcataataattgtctcaac 31

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.on.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	31	6	BD082353 76 kDa, 3
2	31	100.0	31	6	BD082355 76 kDa, 3
C 3	31	100.0	1448	6	BD082347 76 kDa, 3
C 4	31	100.0	2825	6	BD009865 Proteins,
5	31	100.0	10532	1	AE000826 Helicobac
6	31	100.0	15292	1	AE001541 Helicobac
C 7	29.4	94.8	968	6	BD061867 Antigenic
C 8	29.4	94.8	970	6	BD061998 Antigenic
9	25.4	81.9	30	6	AX793134 Sequence
10	23.2	74.8	184688	2	AC102439 Mus muscu
C 11	23.2	74.8	212995	10	AC107741 Mus muscu
C 12	22.8	73.5	120794	2	AC017240 Drosophil
C 13	22.8	73.5	166863	3	AC009741 Drosophil
C 14	22.8	73.5	225653	3	AE003695 Drosophil
C 15	22.6	72.9	110000	1	Continuation (7 of
C 16	22.6	72.9	269050	1	AL935255 Lactobaci
C 17	22.2	71.6	175477	5	AL954814 Zebrafish
C 18	22.2	71.6	189773	2	CR376780 Danio rer
19	22.2	71.6	217717	2	CR293523 Danio rer

20	22	71.0	2532	5	AF081861 Richardso
C 21	22	71.0	42627	3	AY190944 Drosophil
C 22	22	71.0	104995	9	AC136189 Homo sapi
C 23	22	71.0	110000	1	Continuation (6 of
C 24	22	71.0	127922	9	AC134998 Homo sapi
C 25	22	71.0	142250	10	AC132329 Mus muscu
C 26	22	71.0	168774	2	AC046177 Homo sapi
C 27	22	71.0	197208	10	AC122826 Mus muscu
C 28	22	71.0	299050	1	AP003131 Staphyloc
C 29	22	71.0	325350	1	AP004823 Staphyloc
C 30	22	71.0	343590	1	AP003359 Staphyloc
C 31	22	71.0	347572	2	AC107303 Homo sapi
C 32	21.8	70.3	157216	5	EX897661 Zebrafish
C 33	21.8	70.3	159610	5	EX088559 Zebrafish
C 34	21.8	70.3	160487	5	EX005126 Zebrafish
C 35	21.8	70.3	162785	5	EX547938 Zebrafish
C 36	21.8	70.3	166133	9	AC092361 Homo sapi
C 37	21.8	70.3	166189	2	EX470264 Danio rer
C 38	21.8	70.3	170524	2	AC138921 Homo sapi
C 39	21.8	70.3	175762	2	EX957304 Danio rer
C 40	21.8	70.3	178408	2	CR450755 Danio rer
C 41	21.8	70.3	184142	9	AC138828 Homo sapi
C 42	21.8	70.3	189043	2	AC139804 Homo sapi
C 43	21.8	70.3	189471	2	AC139498 Homo sapi
C 44	21.8	70.3	190095	2	AC139806 Homo sapi
C 45	21.8	70.3	195579	2	AC139810 Homo sapi

ALIGNMENTS

RESULT 1
BD082353
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BD082353 31 bp DNA linear PAT 27-AUG-2002
76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
corresponding polynucleotide molecules.

BD082353.1 GI:22627963
JP 2001523954-A/60.

Mastadenovirus

Viruses; dsDNA viruses, no RNA stage; Adenoviridae.

1 (bases 1 to 31)

Kleanthous, H., Lissolo, L., Tomb, J.F., Miller, C. and Garawi, A.A.

76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and

corresponding polynucleotide molecules

Patent: JP 2001523954-A 60 27-NOV-2001;

MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET

VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,

HUMAN GENOME SCIENCES INC

PN JP 2001523954-A/60

PD 27-NOV-2001

PF 31-MAR-1998 JP 1998541962

PI HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI

MILLER,

PI AMAL AL GARAWI

PC A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,

PC A61K39/40,

PC G01N33/554, G01N33/569

CC Strandedness: Single;

CC Topology: Linear;

FH Key Location/Qualifiers.

FEATURES

source
1..31
/organism="Mastadenovirus"
/mol_type="genomic DNA"
/db_xref="taxon:10509"

ORIGIN

Query Match 100.0%; Score 31; DB 6; Length 31;

Best Local Similarity 100.0%; Pred. No. 3.6; Indels 0;

Matches 31; Conservative 0; Mismatches 0; Gaps 0;

```

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
    |||||||||||||||||||||||||||||||
Db 1 GCTAATATTATTCATAATATTGCTCACAAC 31

RESULT 2
BD082355
LOCUS          31 bp DNA linear PAT 27-AUG-2002
DEFINITION    76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
               corresponding polynucleotide molecules.
ACCESSION    BD082355
VERSION      BD082355.1 GI:22627965
KEYWORDS     JP 2001523954-A/62.
SOURCE       Mastadenovirus
ORGANISM     Mastadenovirus
             Viruses; dsDNA viruses, no RNA stage; Adenoviridae.
REFERENCE    1 (bases 1 to 31)
AUTHORS      Kleanthous,H., Lissolo,L., Tomb,J.F., Miller,C. and Garawi,A.A.
TITLE        76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
             corresponding polynucleotide molecules
JOURNAL      Patent: JP 2001523954-A 62 27-NOV-2001;
             MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET
             VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
             HUMAN GENOME SCIENCES INC
COMMENT      PN JP 2001523954-A/62
             PD 27-NOV-2001
             PF 31-MAR-1998 JP 1998541962
             PI HAROLD KLEANTHOUS,LING LISSOLO,JEAN FRANCOIS TOMB,CHARLES PI
             MILLER,
             PI AMAL AL GARAWI
             PC A01N43/04,A01N59/16,A61K9/48,A61K31/70,A61K31/715,A61K39/02,
             PC A61K39/40,
             PC G01N33/554,G01N33/569
             CC Strandedness: Single;
             CC Topology: Linear;
             FH Key Location/Qualifiers.
FEATURES     source
             1..31
             /organism="Mastadenovirus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10509"
ORIGIN
Query Match 100.0%; Score 31; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
    |||||||||||||||||||||||||||||||
Db 1 GCTAATATTATTCATAATATTGCTCACAAC 31

RESULT 3
BD082347/c
LOCUS          1448 bp DNA linear PAT 27-AUG-2002
DEFINITION    76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
               corresponding polynucleotide molecules.
ACCESSION    BD082347
VERSION      BD082347.1 GI:22627957
KEYWORDS     JP 2001523954-A/54.
SOURCE       Mastadenovirus
ORGANISM     Mastadenovirus
             Viruses; dsDNA viruses, no RNA stage; Adenoviridae.
REFERENCE    1 (bases 1 to 1448)
AUTHORS      Kleanthous,H., Lissolo,L., Tomb,J.F., Miller,C. and Garawi,A.A.
TITLE        76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and
             corresponding polynucleotide molecules
JOURNAL      Patent: JP 2001523954-A 54 27-NOV-2001;
             MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET
             VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
             HUMAN GENOME SCIENCES INC
COMMENT      PN JP 2001523954-A/54
             PD 27-NOV-2001

```

```

PF 31-MAR-1998 JP 1998541962
PI HAROLD KLEANTHOUS,LING LISSOLO,JEAN FRANCOIS TOMB,CHARLES PI
MILLER,
PI AMAL AL GARAWI
PC A01N43/04,A01N59/16,A61K9/48,A61K31/70,A61K31/715,A61K39/02,
PC A61K39/40,
PC G01N33/554,G01N33/569
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES     source
             1..1448
             /organism="Mastadenovirus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10509"
ORIGIN
Query Match 100.0%; Score 31; DB 6; Length 1448;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
    |||||||||||||||||||||||||||||||
Db 1324 GCTAATATTATTCATAATATTGCTCACAAC 1294

RESULT 4
BD009865/c
LOCUS          2825 bp DNA linear PAT 31-JAN-2002
DEFINITION    Proteins, in particular membrane proteins, of Helicobacter pylori,
               their preparation and use.
ACCESSION    BD009865
VERSION      BD009865.1 GI:18639238
KEYWORDS     JP 2001502886-A/6.
SOURCE       Helicobacter pylori
ORGANISM     Helicobacter pylori
             Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
             Helicobacteraceae; Helicobacter.
REFERENCE    1 (bases 1 to 2825)
AUTHORS      Knapp,B., Hundt,E. and Schmidt,K.H.
TITLE        Proteins, in particular membrane proteins, of Helicobacter pylori,
             their preparation and use
JOURNAL      Patent: JP 2001502886-A 6 06-MAR-2001;
             CHIRON BEHRING GMBH & CO
COMMENT      OS Helicobacter pylori
             PN JP 2001502886-A/6
             PD 06-MAR-2001
             PF 25-JUL-1997 JP 1998508651
             PF 26-JUL-1996 DE 196 30 390.7
             PI BERNHARD KNAPP,ERIKA HUNDT,KARL HEINZ SCHMIDT PC
             C12N15/31,C07K14/205,C07K16/12,G01N33/53,A61K31/70,A61K39/106, PC
             A61K39/395
             CC Strandedness: Single;
             CC Topology: Linear;
             FH Key Location/Qualifiers
             FT source 1..2825
             /organism="Helicobacter pylori"
             /mol_type="genomic DNA"
             /db_xref="taxon:210"
FEATURES     source
             1..2825
             /organism="Helicobacter pylori"
             /mol_type="genomic DNA"
             /db_xref="taxon:210"
ORIGIN
Query Match 100.0%; Score 31; DB 6; Length 2825;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
    |||||||||||||||||||||||||||||||
Db 2097 GCTAATATTATTCATAATATTGCTCACAAC 2067

```

RESULT 5
 LOCUS AE000626 10532 bp DNA linear BCT 06-APR-1999
 DEFINITION Helicobacter pylori 26695 section 104 of 134 of the complete genome.
 ACCESSION AE000626 AE000511
 VERSION AE000626.1 GI:2314360
 KEYWORDS
 SOURCE Helicobacter pylori 26695
 ORGANISM Helicobacter pylori 26695
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 Helicobacteraceae; Helicobacter.
 REFERENCE 1 (bases 1 to 10532)
 AUTHORS Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Matthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.
 TITLE The complete genome sequence of the gastric pathogen Helicobacter pylori
 JOURNAL Nature 388 (6642), 539-547 (1997)
 MEDLINE 97394467
 PUBMED 9252185
 REFERENCE 2 (bases 1 to 10532)
 AUTHORS Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Matthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 REFERENCE 3 (bases 1 to 10532)
 AUTHORS White, O.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1. 10532
 /organism="Helicobacter pylori 26695"
 /mol_type="genomic DNA"
 /strain="26695"
 /db_xref="taxon:85962"
 complement(142..519)
 /gene="HP1199"
 complement(142..519)
 /gene="HP1199"
 /note="similar to SP:P05392 percent identity: 65.04; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L7/L12 (rpl7/l12)"
 /protein_id="AAD08245.1"
 /db_xref="GI:2314361"
 /translation="MAISKEEVLEIGLSVLELSLVKMFKEKFGVSATPTVVAGAA VAGAAAESEKTFNVLADSGAEKIKVIVREITGLGLKEAKDATEKTPHVLKEG VNKEAEITKKLEVGAKVEVK"
 complement(565..1059)
 /gene="HP1200"
 complement(565..1059)
 /gene="HP1200"
 /note="similar to SP:P42923 PID:786163 GB:AL009126 percent identity: 30.38; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1

gene
 CDS
 /product="ribosomal protein L10 (rpl10)"
 /protein_id="AAD08246.1"
 /db_xref="GI:2314362"
 /translation="MQHQHQHQHVLELVANLKSQFADAKALLICDYKGLSVRKLEALR NKARNGIKQVQIKNTLAHIAKMTGYSDLDLKTNTVFLWGGDOIALLSKLVDFDFOKEH KDHEVLKAGLFDKESVSVHAEVSKLPSEKELMGMLLSVWTAPARYPTVGLDNLKKA KEEN"
 complement(1168..1872)
 /gene="HP1201"
 complement(1168..1872)
 /gene="HP1201"
 /note="similar to GB:L42023 SP:P44342 PID:1003916
 PID:122450 PID:1204767 percent identity: 51.98; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L11 (rpl11)"
 /protein_id="AAD08247.1"
 /db_xref="GI:2314363"
 /translation="MAKKVFKRLEKLFKSIQNDKAYQVEQGVVVKSLASAKFDETVE VALRGVDPHADQMVRAQVVLPHGTGKGVRAVPAKDIKQDEAKNAGADVGGDLA EEIKNGRIDFDMVIATPMMAVGVGRILGPKGLMPNPKTGTVTMDIAKAVSNVAKSG QVNFVRDKGNVHAPIGKASFPPEEKIKENMLELVKTINRLKPSAKGKYIRNAALSLT MSPSVSLDAQELMDIK"
 complement(1917..2342)
 /gene="HP1202"
 complement(1917..2342)
 /gene="HP1202"
 /note="similar to SP:P29395 GB:Z11839 PID:48184
 GB:AE000512 percent identity: 63.83; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L11 (rpl11)"
 /protein_id="AAD08248.1"
 /db_xref="GI:2314364"
 /translation="MAKKVGEIKLQIPAGKANSPVGPALGQGVNIMEFCKAFNE KQDMGSFNPIVITVYQDSFTFKPKPVTDLIKASGVKSGSDNPKNKIAKLTH KQVEEIAQLKMEDLNTSTMEAAKKIVMGSAKSMGVEVVD"
 complement(2360..2890)
 /gene="HP1203"
 complement(2360..2890)
 /gene="HP1203"
 /note="similar to SP:P16921 PID:147801 PID:396321
 GB:U00096 PID:1790414 percent identity: 41.04; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="transcription termination factor NusG (nusG)"
 /protein_id="AAD08249.1"
 /db_xref="GI:2314365"
 /translation="MMDWYAIQYSGSEQSVKKALENLNDHNRIDRIQETIIVPTEDI IEVSKSKTKVTERSLYGVYPIKVDLDVLVHKKIOSLPVSRREIGENKKPTPLSEAD IGHLEKNNRRAAPKPKIPFEQGVVRVVEGFANFTATVEEYDVEHRKLLANVSIFG RNTPIELHLSQVEKII"
 complement(3242..3314)
 /gene="trna-trp-1"
 complement(3242..3314)
 /gene="trna-trp-1"
 /product="trna-trp-1"
 /note="codon recognized: UGG"
 complement(3354..3512)
 /gene="HP1204"
 complement(3354..3512)
 /gene="HP1204"
 /note="similar to SP:P23375 percent identity: 55.10; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L33 (rpl33)"
 /protein_id="AAD08255.1"
 /db_xref="GI:2314371"
 /translation="MKVKIGLKSCDCEDINYSVTTKNAKTNTKLEKFKCPRENKHTL

gene
HKEIKLKS"
/complement (3548. .4747)
/gene="HP1205"
CDS
/complement (3548. .4747)
/gene="HP1205"
/note="similar to SP:P42482 PID:587590 percent identity:
89.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="translation elongation factor EF-Tu (tufB)"
/protein_id="AAD08250.1"
/db_xref="GI:2314366"
/translation="MAKEFNRTKPHVNIPTIGHVDHGGTTLSAIAISAVLSLKLGAEM
KDYDNIDNAPEKERTIATSHLEYETENBHYAHVDCPGHADYVKNMITGAQMDGA
ILVVSADGPMPTREHLLSRQGVPHIVVFLNKQDMDVDDQELLELMEMEVELLSA
YEPFGDDTPIVAGSALRALEAKAGNVGEGKVLKMAEVDYIPTPERDTEKFLM
PVEDVFSIAGTGVTVTRIGRVKGVDEIVGIRPTQKTVTVGVNFRKELEKGEA
GDNVGLLRGTTKEEVERGWLCKPGSITPHKKFEGEIVYLSKBEGRHTPPFTNRP
QFYVYRTDVTGSITLPEGVEMVMPGDNVKITVELISPALEBLGTFKFAIREGRTVGAG
VVSNIIE"
complement (4877. .4948)
/gene="CRNA-Thr-1"
tRNA
complement (4877. .4948)
/gene="CRNA-Thr-1"
/product="CRNA-Thr"
/note="codon recognized: ACC"
complement (4982. .5055)
/gene="CRNA-Gly-2"
gene
complement (4982. .5055)
/gene="CRNA-Gly-2"
tRNA
/product="CRNA-Gly"
/note="codon recognized: GGA"
complement (5074. .5155)
/gene="CRNA-Tyr-1"
gene
complement (5074. .5155)
/gene="CRNA-Tyr-1"
tRNA
/product="CRNA-Tyr"
/note="codon recognized: UAC"
complement (5186. .5258)
/gene="CRNA-Thr-2"
tRNA
complement (5186. .5258)
/gene="CRNA-Thr-2"
/product="CRNA-Thr"
/note="codon recognized: ACA"
5448. .7184
gene
/gene="HP1206"
CDS
5448. .7184
/gene="HP1206"
/note="similar to SP:P22638 PID:142020 percent identity:
26.18; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="multidrug resistance protein (heta)"
/protein_id="AAD08251.1"
/db_xref="GI:2314367"
/translation="MAKKKHISTLKYFLRSIKQIYMLITFKERXWFFLLVMAVFS

Query Match 100.0%; Score 31; DB 1; Length 10532;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCTAATATATTCATAATATTTGCTCAAC 31
Db 3541 GCTAATATATTCATAATATTTGCTCAAC 3571
|||||
|||

RESULT 6
AE001541
LOCUS AE001541 15292 bp DNA linear BCT 20-JAN-1999
DEFINITION Helicobacter pylori, strain J99 section 102 of the complete
genome.
ACCESSION AE001541 AE001439
VERSION AE001541.1 GI:4155724

KEYWORDS
SOURCE
ORGANISM
Helicobacter pylori J99
Helicobacter pylori J99
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
Helicobacteraceae; Helicobacter.
REFERENCE
AUTHORS
1 (bases 1 to 15292)
Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C.,
Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G.,
Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C.,
Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E.,
Vovis, G.F. and Trust, T.J.
Genomic-sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori
99120557
9923682
2 (bases 1 to 15292)
King, B.L., Alm, R.A. and Trust, T.J.
Direct Submission
Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
Street, Cambridge, MA 02139, USA
Address all correspondence to: hp@arch.us.astro.com or Richard
A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
Corporation, 100 Beaver Street, Waltham, MA 02453. Qin Jiang and
Diane E. Taylor are with the University of Alberta Department of
Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G
2H7 and the Canadian Bacterial Diseases Network. All other authors
are with Astra Research Center Boston, 128 Sidney Street,
Cambridge, MA, 02139. Putative identifications, sequence
alignments, and name and sequence search capability are available
at ARCB's World Wide Web site. (URL:
http://www.astro-boston.com/hpylori).
FEATURES
source
1. 15292
/organism="Helicobacter pylori J99"
/mol_type="genomic DNA"
/strain="J99"
/db_xref="taxon:85963"
complement (257. .415)
/gene="rpmG"
/note="synonym: jhp1127"
complement (257. .415)
/gene="rpmG"
/note="similar to H. pylori 26695 gene HP1204"
/codon_start=1
/transl_table=11
/product="50S RIBOSOMAL PROTEIN L33"
/protein_id="AAD06710.1"
/db_xref="GI:4155730"
/translation="MKVKGIGKSCDEDINVTYTKNAKTNTKLELKKFKCPRENKHTL
HKEIKLKS"
complement (462. .1661)
/gene="tufA"
/note="synonym: jhp1128"
complement (462. .1661)
/gene="tufA"
/note="similar to H. pylori 26695 gene HP1205"
/codon_start=1
/transl_table=11
/product="ELONGATION FACTOR TU (EF-TU)"
/protein_id="AAD06711.1"
/db_xref="GI:4155731"
/translation="MAKEFNRTNPHVNIPTIGHVHGKTTLSAIAISAVLSLKLGAEM
KDYDNIDNAPEKERTIATSHLEYETENBHYAHVDCPGHADYVKNMITGAQMDGA
ILVVSADGPMPTREHLLSRQGVPHIVVFLNKQDMDVDDQELLELMEMEVELLSA
YEPFGDDTPIVAGSALRALEAKAGNVGEGKVLKMAEVDYIPTPERDTEKFLM
PVEDVFSIAGTGVTVTRIGRVKGVDEIVGIRPTQKTVTVGVNFRKELEKGEA
GDNVGLLRGTTKEEVERGWLCKPGSITPHKKFEGEIVYLSKBEGRHTPPFTNRP
QFYVYRTDVTGSITLPEGVEMVMPGDNVKITVELISPALEBLGTFKFAIREGRTVGAG
VVSNIIE"


```

JOURNAL      Patent: JP 2001517091-A 201 02-OCT-2001;
GENELABS TECHNOLOGIES INC
COMMENT      PN JP 2001517091-A/201
PD 02-OCT-2001
PF 25-APR-1998 JP 1998547263
PR 25-APR-1997 US 60/045107,14-OCT-1997 US 60/061958 PI
THERESA P CHOW,KIRK E FRY,MOON Y LIM,C P MCATEE PC
C12N15/31,C07K14/205,C07K16/12,A61K39/106
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES     Location/Qualifiers
source       1..968
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"
ORIGIN
Query Match      94.8%; Score 29.4; DB 6; Length 968;
Best Local Similarity 96.8%; Pred. No. 5.1;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCAAC 31
   |||||
Db 722 GCTAATATTATTCATAATATTGCTCAAC 692

RESULT 8
BD061998/4
LOCUS      BD061998 970 bp DNA linear PAT 27-AUG-2002
DEFINITION Antigenic composition and method of detection for Helicobacter
            Pylori.
ACCESSION  BD061998
VERSION     BD061998.1 GI:22607603
KEYWORDS    JP 2001517091-A/332.
SOURCE      synthetic construct
ORGANISM    other sequences; artificial sequences.
REFERENCE   1 (bases 1 to 970)
AUTHORS     Chow,T.P., Fry,K.E., Lim,M.Y. and Mcatee,C.P.
TITLE       Antigenic composition and method of detection for Helicobacter
JOURNAL     Patent: JP 2001517091-A 332 02-OCT-2001;
GENELABS TECHNOLOGIES INC
COMMENT      PN JP 2001517091-A/332
PD 02-OCT-2001
PF 25-APR-1998 JP 1998547263
PR 25-APR-1997 US 60/045107,14-OCT-1997 US 60/061958 PI
THERESA P CHOW,KIRK E FRY,MOON Y LIM,C P MCATEE PC
C12N15/31,C07K14/205,C07K16/12,A61K39/106
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES     Location/Qualifiers
source       1..970
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"
ORIGIN
Query Match      94.8%; Score 29.4; DB 6; Length 970;
Best Local Similarity 96.8%; Pred. No. 5.1;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCAAC 31
   |||||
Db 724 GCTAATATTATTCATAATATTGCTCAAC 694

RESULT 9
AX793134
LOCUS      AX793134 30 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequence 5598 from Patent WO02066501.
ACCESSION  AX793134

```

```

AX793134.1 GI:32958581
VERSION     .
KEYWORDS    Helicobacter pylori
SOURCE      Helicobacter pylori
ORGANISM    Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
            Helicobacteraceae; Helicobacter.
REFERENCE   1
AUTHORS     Legrain,P., Rain,J.C., Colland,F., de Reuse,H. and Labigne,A.
TITLE       Protein-protein interactions in Helicobacter pylori
JOURNAL     Patent: WO 02066501-A 5598 29-AUG-2002;
            Hybrigenics (FR) ; INSTITUT PASTEUR (FR)
FEATURES     Location/Qualifiers
source       1..30
            /organism="Helicobacter pylori"
            /mol_type="unassigned DNA"
            /db_xref="taxon:210"
ORIGIN
Query Match      81.9%; Score 25.4; DB 6; Length 30;
Best Local Similarity 96.3%; Pred. No. 2.8e+02;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCATAATATTGCTCAC 28
   |||||
Db 4 CTACTATTATTCATAATATTGCTCAC 30

AC102439 184688 bp DNA linear HTG 24-MAR-2004
MUS MUSCULUS chromosome 14 clone RP24-14314 map 14, *** SEQUENCING
IN PROGRESS ***, 3 unordered pieces.
AC102439 3 GI:45680590
HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
MUS MUSCULUS (house mouse)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 184688)
AUTHORS       Birren,B., Nusbaum,C. and Lander,E.
TITLE         Mus musculus chromosome 14, clone RP24-14314
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 184688)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campapiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Hearford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneses,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            3 (bases 1 to 184688)
            Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,

```


Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-JAN-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 28, 2004 this sequence version replaced gi:38524693.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L20348
Center clone name: 335_P_23

FEATURES

source

Location/Qualifiers

```

1..212995
  /organism="Mus musculus"
  /mol_type="Genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="3"
  /map="3"
  /clone="RP23-335P23"
  /clone_lib="RPCI-23 Female Mouse BAC"
1..20010
  /note="wgs end extension"
  clone_end:SP6"
34..39
  /note="x30 qual SINGL region"
  complement(259..381)
  /rpt_family="MIR"
  complement(430..587)
  /rpt_family="ORR1D"
  complement(930..1069)
  /rpt_family="ORR1D"
  1527..1658
  /rpt_family="CA)n"
  1670..1729
  /rpt_family="CA)n"
  1774..1918
  /rpt_family="CA)n"
  1998..2043
  /rpt_family="CA)n"
  2377..2476
  /rpt_family="MIR3"
  3457..3523
  /rpt_family="T-rich"
  3715..3738
  /rpt_family="T)n"
  complement(4574..4719)
  /rpt_family="B1_MM"
  complement(4796..4886)
  /rpt_family="PB1D10"
  complement(4823..4908)
  /rpt_family="ID_B1"
  complement(4912..4981)
  /rpt_family="ID3"
  5832..6107
  /rpt_family="LIMC2"
  6108..6230
  /rpt_family="CA)n"
  6231..6325
  /rpt_family="LIMC2"
  6398..6420
  /rpt_family="TTG)n"
  complement(6972..7149)

```

```

repeat_region      /rpt_family="Lx4"
7278..7558
  /rpt_family="B4"
7630..7655
  /rpt_family="(A)n"
8955..8992
  /rpt_family="(TTA)n"
complement(8996..9212)
  /rpt_family="ID_B1"
12006..12062
  /note="single clone coverage"
complement(13110..13443)
  /rpt_family="MTC"
13602..13852
  /rpt_family="ORR1D"
13853..14253
  /rpt_family="MTA"
14254..14329
  /rpt_family="ORR1D"
14368..14558
  /rpt_family="MIR"
16132..16240
  /rpt_family="MIR"
16740..16893
  /rpt_family="(TG)n"
17402..17453
  /rpt_family="(TGAG)n"
17760..17950
  /rpt_family="ID_B1"
17964..18160
  /rpt_family="B3"
complement(18498..18792)
  /rpt_family="B4A"
19418..19438
  /rpt_family="AT-rich"
19504..19796
  /rpt_family="B4A"
20011..20016
  /note="clone boundary"
clone_end:SP6"
site:EcoRI"
20450..20737
  /rpt_family="B4A"
20741..20796
  /rpt_family="(CA)n"
21001..21159
  /rpt_family="LIMB1"
21252..21304
  /rpt_family="CT-rich"

```

Query Match 74.8%; Score 23.2; DB 10; Length 212995;
Best Local Similarity 89.3%; Pred.No.1.5e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTCCTCACAA 30
|||||
Db 22532 TAATATTATTCAATAATATTCCTTAAAA 22505

RESULT 12

AC017240

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 120794)

AUTHORS

Adams, M. and Venter, J.C.

AC017240 AC017240 120794 bp DNA linear HTG 09-DEC-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.

AC017240.1 GI:6553746

HTG; HTGS_PHASE2.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 120794)

Adams, M. and Venter, J.C.

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhou, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C. The genome sequence of <i>Drosophila melanogaster</i> Science 287 (5461), 2185-2195 (2000) 20196006 10731132	gene	/db_xref=FLYBASE:FBgn0051356" complement (4317..6517) /locus_tag="CG4830" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037996" complement (join(4317..5471,5561..5677,5747..5933,6046..6253,6325..6517)) /locus_tag="CG4830" /product="CG4830-RA" /db_xref=FLYBASE:FBgn0037996" complement (join(4520..5471,5561..5677,5747..5933,6046..6253,6325..6465)) /locus_tag="CG4830" /codon_start=1 /product="CG4830-PA" /protein_id="AAF54758.1" /db_xref="GI:7299573" /db_xref=FLYBASE:FBgn0037996"
	2 (bases 1 to 225655) Celinker, S.E., Wheeler, D.A., Krommiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., Hodgson, C.A., George, R.A., Hoskins, R.A., Lavery, T., Muzny, D.M., Nelson, C.R., Pacleb, J.M., Park, S., Pfeiffer, B.D., Richards, S., Sodergren, E.J., Svirskas, R., Tabor, P.E., Wan, K., Stapleton, M., Sutton, G.G., Venter, C., Weinstock, G., Scherer, S.E., Myers, E.W., Gibbs, R.A. and Rubin, G.M. Finishing a whole-genome shotgun: release 3 of the <i>Drosophila melanogaster</i> euchromatic genome sequence Genome Biol. 3 (12), RESEARCH0079 (2002) 22426065 12537568	CDS	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15850)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	2 (bases 1 to 225655) Celinker, S.E., Wheeler, D.A., Krommiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., Hodgson, C.A., George, R.A., Hoskins, R.A., Lavery, T., Muzny, D.M., Nelson, C.R., Pacleb, J.M., Park, S., Pfeiffer, B.D., Richards, S., Sodergren, E.J., Svirskas, R., Tabor, P.E., Wan, K., Stapleton, M., Sutton, G.G., Venter, C., Weinstock, G., Scherer, S.E., Myers, E.W., Gibbs, R.A. and Rubin, G.M. Finishing a whole-genome shotgun: release 3 of the <i>Drosophila melanogaster</i> euchromatic genome sequence Genome Biol. 3 (12), RESEARCH0079 (2002) 22426065 12537568	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15850)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	2 (bases 1 to 225655) Celinker, S.E., Wheeler, D.A., Krommiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., Hodgson, C.A., George, R.A., Hoskins, R.A., Lavery, T., Muzny, D.M., Nelson, C.R., Pacleb, J.M., Park, S., Pfeiffer, B.D., Richards, S., Sodergren, E.J., Svirskas, R., Tabor, P.E., Wan, K., Stapleton, M., Sutton, G.G., Venter, C., Weinstock, G., Scherer, S.E., Myers, E.W., Gibbs, R.A. and Rubin, G.M. Finishing a whole-genome shotgun: release 3 of the <i>Drosophila melanogaster</i> euchromatic genome sequence Genome Biol. 3 (12), RESEARCH0079 (2002) 22426065 12537568	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15850)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	3 (bases 1 to 225655) Mishra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradscky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochnik, S.E., Smith, C.D., Tupy, J.L., Whitfield, E.J., Bayraktaroglu, L., Beran, B.P., Bettencourt, B.R., Celinker, S.E., de Grey, A.D., Drysdale, R.A., Harris, N.L., Richter, J., Russo, S., Schroeder, A.J., Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M., Rubin, G.M. and Lewis, S.E. Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a systematic review Genome Biol. 3 (12), RESEARCH0083 (2002) 22426069 12537572	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15850)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	4 (bases 1 to 225655) Kaminker, J.S., Bergman, C.M., Krommiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M. and Celinker, S.E. The transposable elements of the <i>Drosophila melanogaster</i> euchromatin: a genomics perspective Genome Biol. 3 (12), RESEARCH0084 (2002) 22426070 12537573	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	5 (bases 1 to 225655) Adams, M.D., Celinker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J. Direct Submission Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA 6 (bases 1 to 225655) FlyBase Direct Submission Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA 7 (bases 1 to 225655) FlyBase Direct Submission Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA On Sep 18, 2002 this sequence version replaced gi:7299572. Location/Qualifiers 1..225655 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /chromosome="3R" 2735..2806 /locus_tag="CR31356" /note="CR31356-RA" /map="87B2-87B2"	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	6 (bases 1 to 225655) FlyBase Direct Submission Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA 7 (bases 1 to 225655) FlyBase Direct Submission Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA On Sep 18, 2002 this sequence version replaced gi:7299572. Location/Qualifiers 1..225655 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /chromosome="3R" 2735..2806 /locus_tag="CR31356" /note="CR31356-RA" /map="87B2-87B2"	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	3 (bases 1 to 225655) Mishra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradscky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochnik, S.E., Smith, C.D., Tupy, J.L., Whitfield, E.J., Bayraktaroglu, L., Beran, B.P., Bettencourt, B.R., Celinker, S.E., de Grey, A.D., Drysdale, R.A., Harris, N.L., Richter, J., Russo, S., Schroeder, A.J., Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M., Rubin, G.M. and Lewis, S.E. Annotation of the <i>Drosophila melanogaster</i> euchromatic genome: a systematic review Genome Biol. 3 (12), RESEARCH0083 (2002) 22426069 12537572	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15850)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	4 (bases 1 to 225655) Kaminker, J.S., Bergman, C.M., Krommiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M. and Celinker, S.E. The transposable elements of the <i>Drosophila melanogaster</i> euchromatin: a genomics perspective Genome Biol. 3 (12), RESEARCH0084 (2002) 22426070 12537573	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	5 (bases 1 to 225655) Adams, M.D., Celinker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J. Direct Submission Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA 6 (bases 1 to 225655) FlyBase Direct Submission Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA 7 (bases 1 to 225655) FlyBase Direct Submission Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA On Sep 18, 2002 this sequence version replaced gi:7299572. Location/Qualifiers 1..225655 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /chromosome="3R" 2735..2806 /locus_tag="CR31356" /note="CR31356-RA" /map="87B2-87B2"	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDLPLSL AALSDQPKRLANRTKGDATIELCKRFDHSLADI VOELNVLQEQTRAEADKVSLLI EFLRTEAEQIETYLKLELRERALLALNSLALVCELNKLCFCQPSWSRWQ WTDNSAGLGIEMWORI CGLTEKMLSEFLVIVDDVLAGHNCERKLPKVINHEVVLSDF ALMQTLFLEQDEQEQVOSTIRIPSQPLSLQTLYHLQIALNSVVPOTLPKPKVLO AFIORLGLKCHYEGLAHARECTKASQNALQLFDLKLERVFAISREERTLDQDIIH AQONQLRDYIDPFDFFELFAEHTAHVSRASRLQGEIGLVLPFSQGAASQGAASAAH SGNGARKGDSKSKSSAASFPFGMSQEWFR" complement (16037..17623) /locus_tag="CG4860" /map="87B2-87B2"
	6 (bases 1 to 225655) FlyBase Direct Submission Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA 7 (bases 1 to 225655) FlyBase Direct Submission Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA On Sep 18, 2002 this sequence version replaced gi:7299572. Location/Qualifiers 1..225655 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /chromosome="3R" 2735..2806 /locus_tag="CR31356" /note="CR31356-RA" /map="87B2-87B2"	gene	/db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /map="87B2-87B2" /db_xref=FLYBASE:FBgn0037998" complement (join(11815..12113,13041..13706,13766..14823,14892..15574,15631..15764)) /locus_tag="CG4848" /product="CG4848-PA" /codon_start=1 /protein_id="AAF54760.1" /db_xref="GI:7299575" /db_xref=FLYBASE:FBgn0037998" /translacion="MTANLNLNVDTLFEQHSVSEIDEVHKKIQSVVENKEBELRTHVGERYDLQADTIAAMQTSGATLMEQVRHVQANCRSLNQQLLGFQSTANASAKDAA LKERNAGKGLQTYGTGTAQIKTALPELIWTHLNDRFYAATLFTSRHISTGLQL DGQALMQKLPVARKOWEILPFPHVTIKQALTLALEREELIQEMTDCLOGLLLDKS DLSVLKSLNLRSSAFNLQSGSPRRVKDRILASLNVSLVELLDCLLQSLYSL LFSLESCASTCPCPSINRMESSESLVHLLPEIIAGPKPOFVLPQLEQLHRSQLDYF WLDKRNALAAHLQOVFALVTNMQTIQDIKSAANTNGRPDVRLEQLHRSQLDYF ARKYVPLINARVREIRSSWASAKLTYEQVLLIEAGSQSPPIQIWEQSDDL

/protein_id="BAF54761.1"
/db_xref="GI:7299576"
/db_xref="FLYBASE:FBgn0037999"
/translation="MQQLIRVARTIGRSSGAWSAIIAGNNGIACIALSETHQIL
QKTCREFANAEAPKARHDEELYPAEQVRRRLGELGMSVTVREEYGGSLDYQAYV
IGMEEVAGDAVSIIVMGVNNLYLGAVOQHGTEQOKDFLVPYTGGEHIAFYALSEPG
NGSDAGASTTAKLQGSYQINGTKAWISNSKEASGGIVFATVDKSMKHGKITAPLTP
KDVFLSLAKKESKMGKATSTCQLVLESDVHVPKRSRVIGAAAGDGFKIAWQSLDCGRIG
IAAQATGIAAALAEADVSKRVAFGRKLARLQIQKLMADMATRVESIRLLTWRAA
WLKNGLPITKEAAMAKLHASESATFCAHQCIQLGNGYTTDLPAELYRNARVTEI
YEGTSEIQIRIVIANAVLELGE"

Query Match 73.5%; Score 22.8; DB 3; Length 225655;
Best Local Similarity 92.3%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCAATAATATTGCTCA 27
|||||
Db 139749 CTAATATTATTCAATAATATTGCTCA 139724
|||||

RESULT 15
BX571856_06/c
WPCOMMENT

Sequence split into 29 fragments LOCUS BX571856 Accession BX571856

Fragment Name	Begin	End
BX571856_00	1	110000
BX571856_01	100001	210000
BX571856_02	200001	310000
BX571856_03	300001	410000
BX571856_04	400001	510000
BX571856_05	500001	610000
BX571856_06	600001	710000
BX571856_07	700001	810000
BX571856_08	800001	910000
BX571856_09	900001	1010000
BX571856_10	1000001	1110000
BX571856_11	1100001	1210000
BX571856_12	1200001	1310000
BX571856_13	1300001	1410000
BX571856_14	1400001	1510000
BX571856_15	1500001	1610000
BX571856_16	1600001	1710000
BX571856_17	1700001	1810000
BX571856_18	1800001	1910000
BX571856_19	1900001	2010000
BX571856_20	2000001	2110000
BX571856_21	2100001	2210000
BX571856_22	2200001	2310000
BX571856_23	2300001	2410000
BX571856_24	2400001	2510000
BX571856_25	2500001	2610000
BX571856_26	2600001	2710000
BX571856_27	2700001	2810000
BX571856_28	2800001	2902619

Continuation (7 of 29) of BX571856 from base 600001 (BX571856 Staphylococcus aureus subsp

Query Match 72.9%; Score 22.6; DB 1; Length 110000;
Best Local Similarity 86.2%; Pred. No. 2.9e+02;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CTAATATTATTCAATAATATTGCTCACAA 30
|||||
Db 16155 CTTATTTTATTATAACATTGCTCACAA 16127
|||||

Search completed: April 24, 2005, 03:21:50
Job time : 812.69 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 10:03:10 ; Search time 203.683 Seconds
(without alignments)
900.968 Million cell updates/sec

Title: US-10-039-183A-15
Perfect score: 31
Sequence: 1 gctaattatttcataataattgtctcacaac 31

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	AAV07972	Helicobac
2	31	100.0	31	AAV07970	Helicobac
C 3	31	100.0	1448	AAV07964	Helicobac
C 4	29.4	94.8	968	AAV90742	Nucleotid
C 5	29.4	94.8	970	AAV90873	Nucleotid
C 6	25.4	81.9	30	ABX69371	Novel Hel
C 7	24	77.4	1200	AA533681	Helicobac
C 8	24	77.4	1200	ACA34860	Prokaryot
9	21.6	69.7	30	AAV07968	Helicobac
10	21	67.7	2595	ABN79830	Fungal 2B
11	21	67.7	2595	ABN79831	Fungal 2B
C 12	21	67.7	5333	ABZ10049	Haematopo
C 13	21	67.7	5333	ABZ10195	Haematopo
C 14	21	67.7	8333	AA545406	Chemical
C 15	21	67.7	8333	ABL33502	Human imm
C 16	21	67.7	8333	ABK28255	DNA trans
17	20.6	66.5	110000	ADF77343	Lactic ac
18	20.2	65.2	2169	ADQ08665	Ciona int
19	20.2	65.2	3536	AAA79704	Eucalyptu
20	20	64.5	2640	AAH76178	A. thalia

21	20	64.5	2671	4	AAH76176	Aah76176 A. thalia
C 22	20	64.5	3129	13	ADR08103	Adr08103 Full leng
C 23	20	64.5	5892	4	ABL12454	Ab112454 Drosophil
C 24	20	64.5	8700	6	ABL33002	Ab133002 Human imm
C 25	20	64.5	11493	5	AAH20175	Aah20175 Mouse spa
C 26	20	64.5	74788	11	ACN45118	Acn45118 Human gen
C 27	20	64.5	110000	6	ABA92787_3	Continuation (4 of
C 28	20	64.5	110000	6	ABA90521_11	Continuation (12 o
C 29	19.8	63.9	228	6	ABN70726	Abn70726 Streptoco
C 30	19.8	63.9	240	6	ABN71501	Abn71501 Streptoco
31	19.8	63.9	507	2	AAV99381	Aav99381 cDNA enco
32	19.8	63.9	1827	10	ADJ56509	Adj56509 Human cDN
33	19.8	63.9	2153	2	AAQ06168	Aaq06168 Encodes H
34	19.8	63.9	2153	2	AAQ32854	Aaq32854 BMPs. 3/2
35	19.8	63.9	2153	2	AAQ23679	Aaq23679 Encodes H
36	19.8	63.9	2153	2	AAQ41295	Aaq41295 Human BMP
37	19.8	63.9	2153	2	AAQ99262	Aaq99262 DNA encod
38	19.8	63.9	2153	6	ABK64540	Abk64540 Human ben
39	19.8	63.9	2153	10	ACF05922	Acf05922 Human bon
40	19.8	63.9	2153	10	ACA56729	Aca56729 Human eig
41	19.8	63.9	2153	12	ADH11561	Adh11561 Human bon
42	19.8	63.9	2153	12	ADI56525	Adi56525 Human pol
43	19.8	63.9	4019	9	ACH04010	Ach04010 Human cDN
44	19.8	63.9	110000	6	ABN71527_11	Continuation (12 o
45	19.8	63.9	225734	12	ADQ59377	Adq59377 Human can

ALIGNMENTS

RESULT 1

AAV07972
ID AAV07972 standard; DNA; 31 BP.

AC AAV07972;

DT 25-MAR-2003 (revised)

DT 02-FEB-1999 (first entry)

DE Helicobacter pylori polypeptide GHPO 750 3' DNA primer.

KW GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;
KW primer; ss.

XX Synthetic.

OS Helicobacter pylori.

XX Helicobacter pylori.

FN WO9843479-A1.

XX PD 08-OCT-1998.

XX PF 31-MAR-1998; 98WO-US006421.

XX PR 01-APR-1997; 97US-00831310.

XX PR 01-APR-1997; 97US-00834666.

(INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

(HUMA-) HUMAN GENOME SCI INC.

PI Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.

XX New isolated Helicobacter polynucleotides - used to develop products for

the diagnosis, prevention and treatment of Helicobacter infections and

gastrointestinal diseases.

PS Claim 5; Page 156; 184pp; English.

XX This 3' primer is used with a 5' primer (see AAV07971) in the PCR

amplification of Helicobacter, e.g. Helicobacter pylori, genomic DNA in

order to obtain DNA (see AAV07964) encoding the a 50 kDa polypeptide (see

AAW73035) designated GHPO 750. The isolated polynucleotide, and encoded

CC polypeptide, can be used to develop vaccines for the treatment and
CC prevention of Helicobacter infections. (Updated on 25-MAR-2003 to correct
CC PI field.)
XX

SQ Sequence 31 BP; 12 A; 6 C; 2 G; 11 T; 0 U; 0 Other;
Query Match 100.0%; Score 31; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAATATTTCATTAATATTGCTCAAC 31
|||||
DB 1 GCTAATATTTCATTAATATTGCTCAAC 31

RESULT 2
ID AAV07970 standard; DNA; 31 BP.
XX

AC AAV07970;

XX 25-MAR-2003 (revised)
DT 02-FEB-1999 (first entry)

XX Helicobacter pylori polypeptide GHPO 1360 5' DNA primer.

DE GHPO 1360; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;
KW primer; ss.

XX Synthetic.
OS Helicobacter pylori.

XX WO9843479-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

XX 01-APR-1997; 97US-00831310.

PR 01-APR-1997; 97US-00834666.

XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.

XX New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastroduodenal diseases.

XX Claim 5; Page 156; 184pp; English.

XX This 5' primer is used with a 3' primer (see AAV07970) in the PCR
CC amplification of Helicobacter, e.g. Helicobacter pylori, genomic DNA in
CC order to obtain DNA (see AAV07963) encoding the mature form of a 32 kDa
CC polypeptide (see AAW73034) designated GHPO 1360. The isolated
CC polynucleotide, and encoded polypeptide, can be used to develop vaccines
CC for the treatment and prevention of Helicobacter infections. (Updated on
CC 25-MAR-2003 to correct PI field.)
XX

SQ Sequence 31 BP; 12 A; 6 C; 2 G; 11 T; 0 U; 0 Other;

Query Match 100.0%; Score 31; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAATATTTCATTAATATTGCTCAAC 31
|||||
DB 1 GCTAATATTTCATTAATATTGCTCAAC 31

RESULT 3

ID AAV07964/c

XX AAV07964 standard; DNA; 1448 BP.

XX AC AAV07964;

XX 25-MAR-2003 (revised)

DT 02-FEB-1999 (first entry)

XX Helicobacter pylori 50 kDa polypeptide GHPO 750 DNA.

XX GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; ss.

XX Helicobacter pylori.

XX Location/Qualifiers

FT 118..1317
CDS /*tag= a

XX WO9843479-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

XX 01-APR-1997; 97US-00831310.

PR 01-APR-1997; 97US-00834666.

XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.

DR P-PSDB; AAW73035.

XX New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastroduodenal diseases.

XX Claim 1; Page 150-152; 184pp; English.

XX This DNA sequence codes for a 50 kDa Helicobacter pylori polypeptide (see
CC AAW73035) designated GHPO 750. A polynucleotide encoding GHPO 750 can be
CC obtained from genomic DNA by PCR amplification (see AAV07971-72. The
CC invention provides polynucleotides (see AAV72001, AAV07912-21 and
CC AAV07963-64) encoding a family 76 kDa Helicobacter polypeptides (see
CC AAW73022-32), GHPO 750 and a 32 kDa polypeptide (see AAW73034). These
CC polynucleotides were initially identified in a search of H. pylori
CC genomic databases. DNA cassettes for expression of the Helicobacter
CC proteins (unprocessed or mature forms) in prokaryotic or eukaryotic cells
CC are provided. The polynucleotides can be used in vaccines to prevent or
CC treat Hb infection in a mammal. Viral (especially poxvirus) or bacterial
CC vectors are used. Products and methods of the invention allow treatment
CC and prevention of gastroduodenal diseases associated with Hb infections,
CC including acute, chronic, and atrophic gastritis, and peptic ulcer
CC diseases, e.g. gastric and duodenal ulcers. Diagnostic and detection
CC methods are also provided. GHPO 750 was demonstrated to be a protective
CC antigen. (Updated on 25-MAR-2003 to correct PI field.)
XX

SQ Sequence 1448 BP; 466 A; 237 C; 364 G; 381 T; 0 U; 0 Other;

Query Match 100.0%; Score 31; DB 2; Length 1448;

Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAATATTTCATTAATATTGCTCAAC 31
|||||
DB 1324 GCTAATATTTCATTAATATTGCTCAAC 1294

RESULT 4

AAV90742/c

ID AAV90742 standard; DNA; 968 BP.
 XX
 AC AAV90742;
 XX
 DT 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX
 DE Nucleotide sequence of clone Z14.ASM from cluster 30.
 XX
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9849314-A2.
 XX
 PD 05-NOV-1998.
 XX
 PF 25-APR-1998; 98WO-US008487.
 XX
 PR 25-APR-1997; 97US-0045107P.
 PR 14-OCT-1997; 97US-0061958P.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 XX
 PI Chow TP, Fry KE, Lim MY, Mcatee CP;
 XX WPI; 1999-009433/01.
 XX
 PS New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX
 PS Claim 27; Page 170; 402pp; English.
 XX
 CC The present sequence encodes a Helicobacter pylori antigenic protein that
 CC is characterised by immunoreactivity with H. pylori-positive antisera.
 CC The proteins are highly immunogenic and induce a long-lasting immune
 CC response that persists even after antimicrobial treatment. In antibody-
 CC detection assays, on sera, plasma, urine, saliva etc., they are highly
 CC sensitive and specific. The specification also describes 69 previously
 CC unrecognised immunogenic cluster families. H. pylori antigens are used to
 CC detect H. pylori-specific antibodies, for diagnosing infection or to
 CC confirm eradication of infection, and in vaccines to protect against H.
 CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
 CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 968 BP; 306 A; 154 C; 260 G; 248 T; 0 U; 0 Other;
 XX
 Query Match 94.8%; Score 29.4; DB 2; Length 968;
 Best Local Similarity 96.8%; Pred. No. 0.42;
 Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 XX
 QY 1 GCTATATATTCATATATATGCTCACAC 31
 Db 722 GCTATATATTCATATATGCTCACAC 692
 XX
 RESULT 5
 AAV90873/c
 ID AAV90873 standard; DNA; 970 BP.
 XX
 AC AAV90873;
 XX
 DT 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX
 DE Nucleotide sequence of cluster 30.
 XX
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 XX
 OS Helicobacter pylori.

XX WO9849314-A2.
 XX
 PD 05-NOV-1998.
 XX
 PF 25-APR-1998; 98WO-US008487.
 XX
 PR 25-APR-1997; 97US-0045107P.
 PR 14-OCT-1997; 97US-0061958P.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 XX
 PI Chow TP, Fry KE, Lim MY, Mcatee CP;
 XX WPI; 1999-009433/01.
 XX
 PS New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX
 PS Claim 22; Page 296; 402pp; English.
 XX
 CC The present sequence encodes a Helicobacter pylori antigenic protein that
 CC is characterised by immunoreactivity with H. pylori-positive antisera.
 CC The proteins are highly immunogenic and induce a long-lasting immune
 CC response that persists even after antimicrobial treatment. In antibody-
 CC detection assays, on sera, plasma, urine, saliva etc., they are highly
 CC sensitive and specific. The specification also describes 69 previously
 CC unrecognised immunogenic cluster families. H. pylori antigens are used to
 CC detect H. pylori-specific antibodies, for diagnosing infection or to
 CC confirm eradication of infection, and in vaccines to protect against H.
 CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
 CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 970 BP; 306 A; 155 C; 261 G; 248 T; 0 U; 0 Other;
 XX
 Query Match 94.8%; Score 29.4; DB 2; Length 970;
 Best Local Similarity 96.8%; Pred. No. 0.42;
 Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 XX
 QY 1 GCTATATATTCATATATATGCTCACAC 31
 Db 724 GCTATATATTCATATATGCTCACAC 694
 XX
 RESULT 6
 ABX69371
 ID ABX69371 standard; DNA; 30 BP.
 XX
 AC ABX69371;
 XX
 DT 07-MAY-2003 (first entry)
 XX
 DE Novel Helicobacter pylori gene PCR primer #2342.
 XX
 KW Protein-protein interaction; ulcer; selected interacting domain; SID;
 KW PCR; primer; ss.
 XX
 OS Helicobacter pylori.
 XX
 PN WO200266501-A2.
 XX
 PD 29-AUG-2002.
 XX
 PF 28-DEC-2001; 2001WO-EP015428.
 XX
 PR 02-JAN-2001; 2001US-0259302P.
 XX
 PA (HYBR-) HYBRIGENICS.
 PA (INSP) INST PASTEUR.
 XX
 PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
 XX

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1200 BP; 365 A; 207 C; 319 G; 309 T; 0 U; 0 Other;

Query Match 77.4%; Score 24; DB 8; Length 1200;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 TTATTCAATAATATTGCTCACAC 31
 |||||
 DB 1200 TTATTCAATAATATTGCTCACAC 1177

RESULT 9
 AAV07968

ID AAV07968 standard; DNA; 30 BP.

XX AAV07968;

XX 25-MAR-2003 (revised)

DT 02-FEB-1999 (first entry)

DE Helicobacter pylori polypeptide GHPO 750 3' DNA primer.

XX GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;
 KW primer; ss.

XX Synthetic.

OS Helicobacter pylori.

XX WO9843479-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

XX 01-APR-1997; 97US-00831310.

PR 01-APR-1997; 97US-00834666.

XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PA (HUNA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Liesolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.

XX New isolated Helicobacter polynucleotides - used to develop products for
 PT the diagnosis, prevention and treatment of Helicobacter infections and

PT gastroduodenal diseases.

XX Example 3.B; Page 64; 194pp; English.

XX This 3' primer was used with a 5' primer (see AAV07967) in the PCR
 CC amplification of Helicobacter pylori strain ORV2001 genomic DNA in order
 CC to obtain DNA (see AAV07964) encoding a 50 kDa polypeptide (see AAV73035)
 CC designated GHPO 750. The primer pair includes a 5' clamp and BamHI and
 CC XhoI restriction enzyme recognition sequences for cloning purposes. The
 CC PCR product was ligated into vector pET28a, and recombinant polypeptide
 CC was expressed as a histidine-tagged fusion protein in *E. coli* host cells.
 CC The polypeptide can be used to develop vaccines for the treatment and
 CC prevention of Helicobacter infections. (Updated on 25-MAR-2003 to correct
 CC PI field.)

SQ Sequence 30 BP; 8 A; 8 C; 4 G; 10 T; 0 U; 0 Other;

Query Match 69.7%; Score 21.6; DB 2; Length 30;

Best Local Similarity 85.7%; Pred. No. 2.4e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GCTAATATTATTCAATAATATTGCTCAC 28

||| |||||
 DB 3 GCTCAGTATTCAATAATATTGCTCAC 30

RESULT 10

ABN79830

ID ABN79830 standard; DNA; 2595 BP.

XX ABN79830;

XX 24-JUL-2002 (first entry)

XX Fungal ZBC gene sequence #67.

XX Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;
 KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;
 KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;
 KW angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;
 KW fungal toxin; cell surface receptor; plant growth regulator; pigment;
 KW insecticide; antineoplastic; gene; ds.

XX Unidentified.

XX WO200224865-A2.

XX 28-MAR-2002.

XX 19-SEP-2001; 2001WO-US029288.

XX 19-SEP-2000; 2000US-0233564P.

XX (MICR-) MICROBIA INC.

XX Holtzman D, Madden K, Maxon M, Sherman A;

XX WPI; 2002-352005/38.

XX P-PSDB; ABP35641.

XX New method for improving the production of a secondary metabolite e.g.
 PT antineoplastic agent, ergot alkaloid from a fungus involves modulation of
 PT the expression of at least one zinc binuclear cluster protein gene.

XX Claim 2; SEQ ID NO 145; 49pp + Sequence Listing; English.

XX The invention relates to improving the production of a secondary
 CC metabolite by a fungus. This involves modulating the expression of at
 CC least one ZBC (zinc binuclear cluster protein) gene in a manner to
 CC improve the yield of the secondary metabolite. Methods of the invention
 CC may be used for improving the production of the secondary metabolite e.g.
 CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such
 CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),

CC an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such
CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,
CC a fungal toxin, a modulator of cell surface receptor signalling, a plant
CC growth regulator, a pigment, an insecticide, or an antineoplastic
CC compound. The method results in a decrease in fermentor run-time, a
CC decrease in the size of the fermentor required for the production of
CC equivalent amounts of the secondary metabolite, or a decrease in the
CC biomass required for the production, which translates into decreased
CC waste that must be handled in downstream processing. The sequences given
CC in records ABN79764-ABN79911 represent ZBC genes of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2595 BP; 829 A; 588 G; 491 G; 687 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 6; Length 2595;
Best Local Similarity 82.8%; Pred. No. 3.8e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTAATATTATTCAATAATATTGCTCACAA 30
||||| ||| ||||||||| |||||
DB 1027 CTAATACTAAACAATAATATTACTCATAA 1055

RESULT 11

ABN79831
ID ABN79831 standard; DNA; 2595 BP.

XX
AC ABN79831;

XX 24-JUL-2002 (first entry)

XX Fungal ZBC gene sequence #68.

XX Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;
KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;
KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;
KW angiogenesis inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;
KW fungal toxin; cell surface receptor; plant growth regulator; pigment;
KW insecticide; antineoplastic; gene; ds.

XX Unidentified.

XX WO200224865-A2.

XX 28-MAR-2002.

XX 19-SEP-2001; 2001WO-US029288.

XX 19-SEP-2000; 2000US-0233564P.

XX (MICR-) MICROBIA INC.

XX Holtzman D, Madden K, Maxon M, Sherman A;

XX WPI; 2002-352005/38.

DR P-PSDB; ABP35642.

XX New method for improving the production of a secondary metabolite e.g.
PT antineoplastic agent, ergot alkaloid from a fungus involves modulation of
PT the expression of at least one zinc binuclear cluster protein gene.

XX Claim 2; SEQ ID NO 147; 49pp + Sequence Listing; English.

XX The invention relates to improving the production of a secondary
CC metabolite by a fungus. This involves modulating the expression of at
CC least one ZBC (zinc binuclear cluster protein) gene in a manner to
CC improve the yield of the secondary metabolite. Methods of the invention
CC may be used for improving the production of the secondary metabolite e.g.
CC antibacterial (such as beta-lactam), an anti-hypercholesterolaemic (such
CC as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),
CC an ergot alkaloid (such as ergotamine), an angiogenesis inhibitor (such

CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,
CC a fungal toxin, a modulator of cell surface receptor signalling, a plant
CC growth regulator, a pigment, an insecticide, or an antineoplastic
CC compound. The method results in a decrease in fermentor run-time, a
CC decrease in the size of the fermentor required for the production of
CC equivalent amounts of the secondary metabolite, or a decrease in the
CC biomass required for the production, which translates into decreased
CC waste that must be handled in downstream processing. The sequences given
CC in records ABN79764-ABN79911 represent ZBC genes of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 2595 BP; 829 A; 588 G; 490 G; 688 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 6; Length 2595;
Best Local Similarity 82.8%; Pred. No. 3.8e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTAATATTATTCAATAATATTGCTCACAA 30
||||| ||| ||||||||| |||||
DB 1027 CTAATACTAAACAATAATATTACTCATAA 1055

RESULT 12

ABZ10049/c

ID ABZ10049 standard; DNA; 5333 BP.

XX
AC ABZ10049;

XX 16-JAN-2003 (first entry)

XX Haematopoietic cell proliferation disorder related DNA sequence #189.

XX Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.

XX Homo sapiens.

XX WO20027272-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WO-EP003401.

XX 26-MAR-2001; 2001US-0278333P.

XX (EPIG-) EPIGENOMICS AG.

XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;

PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;

PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;

PI Schwöpe I, Ziebarth H;

XX WPI; 2003-018942/01.

XX Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.

XX Claim 28; SEQ ID NO 189; 117pp; English.

XX The present invention describes a method for detecting and

CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative

CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclases, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 5333 BP; 1248 A; 87 C; 1254 G; 2744 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 8; Length 5333;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTAATATTATTCATAATATTGCTCACAA 30
 |||||
 DB 2983 CTAATATTATTAATAATATTATTCTTAA 2955

RESULT 13
 ID ABZ10195/c
 XX
 AC ABZ10195;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #335.
 XX
 KW Human; haematopoietic cell proliferation disorder; cytosine;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20027272-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-EP003401.
 XX
 PR 26-MAR-2001; 2001US-0278333P.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model P, Mueller V, Otto T, Pelet C;
 PI Schwöpe I, Ziebarth H;
 XX
 DR WPI; 2003-018942/01.
 XX
 PT Detecting and differentiating between haematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.
 XX
 PS Claim 28; SEQ ID NO 335; 117pp; English.

CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used for
 CC differentiating between healthy haematopoietic cells and proliferative

CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclases, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 5333 BP; 1248 A; 0 C; 1254 G; 2831 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 8; Length 5333;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTAATATTATTCATAATATTGCTCACAA 30
 |||||
 DB 2983 CTAATATTATTAATAATATTATTCTTAA 2955

RESULT 14
 ID AAS45406/c
 XX
 AC AAS45406;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Chemically pretreated genomic DNA associated with cell cycle #56.
 XX
 KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KW PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200168911-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP002945.
 XX
 PR 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602751/68.

PT Designing primers and probes for analyzing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle.
 XX
 PS Claim 1; SEQ ID NO 111; 28pp; English.
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated

CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers
 XX
 SQ Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 4; Length 8333;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTAATATTATTCAATAATATTGCTCACAA 30
 |||||
 Db 5983 CTAATATTATTAAATAATATTATTCTTAA 5955

RESULT 15
 ABL33502/c
 ID ABL33502 standard; DNA; 8333 BP.

XX AC ABL33502;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 1475.

XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.

XX OS Homo sapiens.

XX OS WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP007537.

XX PR 30-JUN-2000; 2000DE-01032529.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130909/17.

XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX PS Claim 1; SEQ ID NO 1475; 32pp + Sequence Listing; German.

XX CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 8333 BP; 1968 A; 183 C; 2037 G; 4145 T; 0 U; 0 Other;

Query Match 67.7%; Score 21; DB 6; Length 8333;
 Best Local Similarity 82.8%; Pred. No. 3.8e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 CTAATATTATTCAATAATATTGCTCACAA 30
 |||||
 Db 5983 CTAATATTATTAAATAATATTATTCTTAA 5955

Search completed: April 24, 2005, 02:20:04
 Job time : 205.683 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:51:24 ; Search time 60.2535 Seconds
(without alignments)
841.853 Million cell updates/sec

Title: US-10-039-183A-15
Perfect score: 31
Sequence: 1 gctaattatttcataataattgtctcacaac 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21.6	69.7	278866	4	US-09-949-016-13922
C 2	21.6	69.7	278866	4	US-09-949-016-13923
C 3	21.6	69.7	278866	4	US-09-949-016-13924
C 4	21.6	69.7	278866	4	US-09-949-016-13925
C 5	21.6	69.7	278866	4	US-09-949-016-13926
C 6	21.6	69.7	278866	4	US-09-949-016-13927
C 7	21.6	69.7	278866	4	US-09-949-016-14699
C 8	21.6	69.7	278866	4	US-09-949-016-14700
C 9	21.6	69.7	278866	4	US-09-949-016-14701
C 10	21.6	69.7	278866	4	US-09-949-016-14702
C 11	20.4	65.8	94077	4	US-09-949-016-14703
C 12	20.2	65.2	47	4	US-09-422-978-317
C 13	20.2	65.2	3536	4	US-10-101-464A-856
C 14	20	64.5	11492	4	US-09-693-205A-3
C 15	20	64.5	43192	4	US-09-949-016-15466
C 16	20	64.5	58789	4	US-09-949-016-15922
C 17	20	64.5	268449	4	US-09-949-016-17244
C 18	20	64.5	640681	4	US-09-790-988-1
C 19	19.8	63.9	507	3	US-08-868-452-41
C 20	19.8	63.9	1866	4	US-09-248-796A-5265
C 21	19.8	63.9	2149	4	US-09-949-016-1611
C 22	19.8	63.9	2153	1	US-07-539-756-1
C 23	19.8	63.9	2153	2	US-08-377-292-5
C 24	19.8	63.9	2153	1	US-07-989-847-9
C 25	19.8	63.9	2153	3	US-08-469-411-9
C 26	19.8	63.9	2153	4	US-09-016-434-1327
C 27	19.8	63.9	2153	4	US-09-780-601A-9

28	19.8	63.9	124110	4	US-09-949-016-13353	Sequence 13353, A
29	19.6	63.2	601	4	US-09-949-016-60540	Sequence 60540, A
30	19.6	63.2	601	4	US-09-949-016-60541	Sequence 60541, A
C 31	19.6	63.2	601	4	US-09-949-016-157403	Sequence 157403, A
C 32	19.6	63.2	601	4	US-09-949-016-157510	Sequence 157510, A
C 33	19.6	63.2	59519	4	US-09-949-016-13504	Sequence 13504, A
C 34	19.6	63.2	126176	4	US-09-949-016-16137	Sequence 16137, A
C 35	19.6	63.2	126176	4	US-09-949-016-16138	Sequence 16138, A
C 36	19.4	62.6	260247	4	US-09-949-016-13358	Sequence 13358, A
C 37	19.2	61.9	23094	4	US-09-949-016-13468	Sequence 13468, A
C 38	19	61.3	282	4	US-09-861-451A-51	Sequence 51, Appl
C 39	19	61.3	601	4	US-09-949-016-170105	Sequence 170105, A
C 40	19	61.3	1409	4	US-09-976-594-349	Sequence 349, App
C 41	19	61.3	2054	4	US-09-949-016-1199	Sequence 1199, App
C 42	19	61.3	2054	4	US-09-949-016-1200	Sequence 1200, App
C 43	19	61.3	2120	4	US-09-949-016-1197	Sequence 1197, App
C 44	19	61.3	2120	4	US-09-949-016-1198	Sequence 1198, App
C 45	19	61.3	2137	4	US-09-976-594-348	Sequence 348, App

ALIGNMENTS

RESULT 1

US-09-949-016-13922/c
; Sequence 13922, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13922
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13922

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TAAATTATTTCATAATATTCCTCACAA 30
|||||
Db 65820 TAAATTATTTCATAATATTCCTCACAA 65793

RESULT 2

US-09-949-016-13923/c
; Sequence 13923, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13923
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13923

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 3
US-09-949-016-13924/c
; Sequence 13924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13924
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13924

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 4
US-09-949-016-13925/c
; Sequence 13925, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13925
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13925

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 5
US-09-949-016-13926/c
; Sequence 13926, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13926
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13926

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
Db 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 6
US-09-949-016-14699/c
; Sequence 14699, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14699
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14699

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
DB 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 7

US-09-949-016-14700/c
; Sequence 14700, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14700
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14700

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
DB 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 8

US-09-949-016-14701/c

; Sequence 14701, Application US/09949016

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14701
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14701

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
DB 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 9

US-09-949-016-14702/c
; Sequence 14702, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14702
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14702

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAATATTATTCAATAATATTGCTCACA 30
|||||
DB 65820 TAATATTATTCAATAATATTGCTCAGAA 65793

RESULT 10

US-09-949-016-14703/c
; Sequence 14703, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14703
; LENGTH: 278866
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(278866)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14703

Query Match 69.7%; Score 21.6; DB 4; Length 278866;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATTAATATTGCTCACA 30
|||||
Db 65820 TAATATTATTCATTAATATTGCTCACA 65793

RESULT 11

US-09-949-016-13635
; Sequence 13635, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13635
; LENGTH: 94077
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13635

Query Match 65.8%; Score 20.4; DB 4; Length 94077;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATTAATATTGCTCACA 30
|||||
Db 45490 GCTATTGTATTCATTAATATTGCTCACA 45519

RESULT 12

US-09-422-978-317/c
; Sequence 317, Application US/09422978
; Patent No. 6537751
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: Biallelic markers for use in constructing a high density...
; FILE REFERENCE: GENSET.020CPI
; CURRENT APPLICATION NUMBER: US/09/422,978
; CURRENT FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: US 09/298,850
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,732
; EARLIER FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: US 60/082,614
; EARLIER FILING DATE: 1998-04-21
; NUMBER OF SEQ ID NOS: 11796
; SEQ ID NO 317
; LENGTH: 47
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 24
; OTHER INFORMATION: 99-14186-424 : polymorphic base A or G
US-09-422-978-317

Query Match 65.2%; Score 20.2; DB 4; Length 47;
Best Local Similarity 81.5%; Pred. No. 90;
Matches 22; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAAATATTATTCATTAATATTGCTCACA 29
|||||
Db 45 TAAATAGTTTCTACATATTGTCACA 19

RESULT 13

US-10-101-464A-856
; Sequence 856, Application US/10101464A
; Patent No. 6768041
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; AND Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 856
; LENGTH: 3536
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-101-464A-856

Query Match 65.2%; Score 20.2; DB 4; Length 3536;
Best Local Similarity 88.0%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 TAAATATTATTCATTAATATTGCTCACA 27
|||||
Db 181 TAAATTTATTCACATATTGCTCA 205

RESULT 14

US-09-693-205A-3/c
; Sequence 3, Application US/09693205A
; Patent No. 6812333
; GENERAL INFORMATION:
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Engert, James C.
; APPLICANT: Richter, Andrea
; TITLE OF INVENTION: IDENTIFICATION OF ARSACS MUTATIONS AND
; FILE REFERENCE: 2825.1021-003
; CURRENT APPLICATION NUMBER: US/09/693,205A
; CURRENT FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/160,588
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 11492
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-693-205A-3

Query Match 64.5%; Score 20; DB 4; Length 11492;
Best Local Similarity 82.1%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATGCTCACA 30
Db 5105 TAATATTGGCATAAATTTTCTCACA 5078

RESULT 15

US-09-949-016-15466/c
; Sequence 15466, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15466
; LENGTH: 43192
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15466

Query Match 64.5%; Score 20; DB 4; Length 43192;
Best Local Similarity 82.1%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATGCTCACA 30
Db 26391 TAAATATTCAATAATATGGAAAAA 26364

Search completed: April 24, 2005, 05:33:08
Job time : 70.4535 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 02:04:39 ; Search time 250.62 Seconds
(without alignments)
750.949 Million cell updates/sec

Title: US-10-039-183A-15

Perfect score: 31

Sequence: 1 gctaattatttcataataattgtctcaaac 31

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 303525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	8	US-08-831-310-15
2	31	100.0	31	8	US-08-831-310-17
3	31	100.0	31	15	US-10-039-183A-15
4	31	100.0	31	15	US-10-039-183A-17
C 5	31	100.0	1448	8	US-08-831-310-3
C 6	31	100.0	1448	15	US-10-039-183A-3
C 7	31	100.0	2825	19	US-10-662-126-32
C 8	24	77.4	1200	9	US-09-815-242-7318
C 9	24	77.4	1200	17	US-10-282-122A-22730
10	21.6	69.7	30	8	US-08-831-310-8
11	21.6	69.7	30	15	US-10-039-183A-8

12	21	67.7	1290	18	US-10-425-115-28120	Sequence 28120, A
13	21	67.7	2595	17	US-10-149-310-145	Sequence 145, App
14	21	67.7	2595	17	US-10-149-310-147	Sequence 147, App
C 15	21	67.7	5333	18	US-10-473-126-189	Sequence 189, App
C 16	21	67.7	5333	18	US-10-473-126-335	Sequence 335, App
C 17	21	67.7	8333	14	US-10-239-676-113	Sequence 113, App
C 18	21	67.7	8333	15	US-10-311-455-1475	Sequence 1475, App
C 19	21	67.7	8333	15	US-10-311-455-129	Sequence 129, App
20	20.6	66.5	1093	17	US-10-425-114-24648	Sequence 24648, A
21	20.6	66.5	1123	17	US-10-425-114-6941	Sequence 6941, App
22	20.6	66.5	1141	17	US-10-425-114-25820	Sequence 25820, A
23	20.6	66.5	1157	17	US-10-425-114-27834	Sequence 27834, A
24	20.6	66.5	1259	18	US-10-425-115-166130	Sequence 166130, A
25	20.6	66.5	1560	18	US-10-425-115-166131	Sequence 166131, A
C 26	20.4	65.8	1803	18	US-10-739-930-22	Sequence 22, Appl
C 27	20.4	65.8	225883	14	US-10-175-523-57	Sequence 57, Appl
C 28	20.2	65.2	47	17	US-10-349-143-317	Sequence 317, Appl
29	20.2	65.2	624	17	US-10-424-599-71415	Sequence 71415, A
30	20.2	65.2	3536	14	US-10-101-464A-856	Sequence 856, App
31	20.2	65.2	3536	19	US-10-864-252-856	Sequence 856, App
32	20	64.5	2640	17	US-10-203-351-3	Sequence 3, Appl
C 33	20	64.5	8700	15	US-10-311-455-975	Sequence 975, App
C 34	20	64.5	74788	13	US-10-087-192-1906	Sequence 1906, App
C 35	20	64.5	87394	18	US-10-810-788A-6	Sequence 6, Appl
36	20	64.5	640681	9	US-09-790-988-1	Sequence 1, Appl
C 37	20	64.5	3673778	16	US-10-312-841-2	Sequence 2, Appl
38	19.8	63.9	1827	15	US-10-084-817-315	Sequence 315, App
39	19.8	63.9	2153	10	US-09-960-706-692	Sequence 692, App
40	19.8	63.9	2153	10	US-09-873-319-435	Sequence 435, App
41	19.8	63.9	2153	15	US-10-286-152A-5	Sequence 5, Appl
42	19.8	63.9	2153	17	US-10-366-345-13	Sequence 13, Appl
43	19.8	63.9	2153	17	US-10-375-150-9	Sequence 9, Appl
44	19.8	63.9	2153	17	US-10-305-720-1327	Sequence 1327, App
45	19.8	63.9	3995	17	US-10-062-674-2020	Sequence 2020, App

ALIGNMENTS

RESULT 1

US-08-831-310-15
; Sequence 15, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831.310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 15:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-831-310-15

Query Match          100.0%; Score 31; DB 8; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATAATTGCTCACAAC 31
    |||||
Db 1 GCTAATATTATTCATAATAATTGCTCACAAC 31

RESULT 2
US-08-831-310-17
; Sequence 17, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-831-310-17

Query Match          100.0%; Score 31; DB 8; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATAATTGCTCACAAC 31
    |||||
Db 1 GCTAATATTATTCATAATAATTGCTCACAAC 31

RESULT 3
US-10-039-183A-15
; Sequence 15, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
```

```
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-15

Query Match          100.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATAATTGCTCACAAC 31
    |||||
Db 1 GCTAATATTATTCATAATAATTGCTCACAAC 31

RESULT 4
US-10-039-183A-17
; Sequence 17, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-17

Query Match          100.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATAATTGCTCACAAC 31
    |||||
Db 1 GCTAATATTATTCATAATAATTGCTCACAAC 31

RESULT 5
US-08-831-310-3/c
; Sequence 3, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
```

```

; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 118...1314
; OTHER INFORMATION:
;
US-08-831-310-3

Query Match 100.0%; Score 31; DB 8; Length 1448;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
Db 1324 GCTAATATTATTCATAATATTGCTCACAAC 1294

RESULT 6
US-10-039-183A-3/c
; Sequence 3, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Liessolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1448
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (118)...(1314)
US-10-039-183A-3

Query Match 100.0%; Score 31; DB 15; Length 1448;
Best Local Similarity 100.0%; Pred. No. 0.33;

US-09-815-242-7318/c
; Sequence 7318, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

US-10-662-126-32/c
; Sequence 32, Application US/10662126
; Publication No. US20050063987A1
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Schmidt, Erika
; APPLICANT: Schmidt, Karl-Heinz
; TITLE OF INVENTION: Proteins, In Particular Membrane Proteins, of
; TITLE OF INVENTION: Helicobacter Pylori, Their Preparation and Use
; FILE REFERENCE: CHIR-0340
; CURRENT APPLICATION NUMBER: US/10/662,126
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: 09/230,158
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: PCT/IB97/00981
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 32
; LENGTH: 2825
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (891)..(2090)
; OTHER INFORMATION: bp protein; 42 kD protein from Helicobacter pylori
US-10-662-126-32

Query Match 100.0%; Score 31; DB 19; Length 2825;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCATAATATTGCTCACAAC 31
Db 2097 GCTAATATTATTCATAATATTGCTCACAAC 2067

RESULT 8
US-09-815-242-7318/c
; Sequence 7318, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7318
 ; LENGTH: 1200
 ; TYPE: DNA
 ; ORGANISM: Helicobacter pylori
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1200)
 US-09-815-242-7318

Query Match 77.4%; Score 24; DB 9; Length 1200;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TTATTCAATAATATTGCTCACAAC 31
 Db 1200 TTATTCAATAATATTGCTCACAAC 1177

RESULT 9
 US-10-282-122A-22730/C
 ; Sequence 22730, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 22730
 ; LENGTH: 1200
 ; TYPE: DNA
 ; ORGANISM: Helicobacter pylori
 US-10-282-122A-22730

Query Match 77.4%; Score 24; DB 17; Length 1200;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TTATTCAATAATATTGCTCACAAC 31
 Db 1200 TTATTCAATAATATTGCTCACAAC 1177

RESULT 10
 US-08-831-310-8
 ; Sequence 8, Application US/08831310
 ; Publication No. US20020026035A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kleanthous, Harold et al.
 ; TITLE OF INVENTION: Helicobacter GHPO 1360 and
 ; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Clark & Elbing LLP
 ; STREET: 176 Federal Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/831,310
 ; FILING DATE: 01-APR-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clark, Paul T.
 ; REGISTRATION NUMBER: 30,175
 ; REFERENCE/DOCKET NUMBER: 06132/037001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-428-0200
 ; TELEFAX: 617-428-7045
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 30 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 US-08-831-310-8

Query Match 69.7%; Score 21.6; DB 8; Length 30;
 Best Local Similarity 85.7%; Pred. No. 3.5e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTAATATTATTCAATAATATTGCTCAC 28
 Db 3 GCTCAGTTATTCAATAATATTGCTCAC 30

RESULT 11
 US-10-039-183A-8
 ; Sequence 8, Application US/10039183A
 ; Publication No. US20030143242A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kleanthous, Harold
 ; APPLICANT: Lissolo, Ling
 ; APPLICANT: Tomb, Jean-Francois
 ; APPLICANT: Miller, Charles
 ; APPLICANT: Al-Garawi, Amal
 ; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
 ; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
 ; FILE REFERENCE: 06132/037002

; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-8

Query Match 69.7%; Score 21.6; -DB 15; Length 30;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCTAATATTATTCAATAATATTGCTCAC 28
DB 3 GCTCAGATTATTCAATAATATTGCTCAC 30

RESULT 12
US-10-425-115-28120
; Sequence 28120, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 28120
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_125658C.1
US-10-425-115-28120

Query Match 67.7%; Score 21; DB 18; Length 1290;
Best Local Similarity 82.8%; Pred. No. 9.7e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCTAATATTATTCAATAATATTGCTCACA 29
DB 1107 GCCAATATTATTCAATAATATTGTCGATA 1135

RESULT 13
US-10-149-310-145
; Sequence 145, Application US/10149310
; Publication No. US20040077039A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; APPLICANT: Maxon, Mary T.
; APPLICANT: Sherman, Amir
; TITLE OF INVENTION: Modulation of Secondary Metabolite Production by
; TITLE OF INVENTION: Zinc Binuclear Cluster Proteins
; FILE REFERENCE: 14184-019US1
; CURRENT APPLICATION NUMBER: US/10/149,310
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: PCT/US01/29288
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 60/233,564
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 145
; LENGTH: 2595
; TYPE: DNA
; ORGANISM: Kluyveromyces lactis
US-10-149-310-145

Query Match 67.7%; Score 21; DB 17; Length 2595;
Best Local Similarity 82.8%; Pred. No. 1.1e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTAATATTATTCAATAATATTGCTCACA 30
DB 1027 CTATACTAAACAATAATATTACTCTAA 1055

RESULT 14
US-10-149-310-147
; Sequence 147, Application US/10149310
; Publication No. US20040077039A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas
; APPLICANT: Madden, Kevin T.
; APPLICANT: Maxon, Mary
; APPLICANT: Sherman, Amir
; TITLE OF INVENTION: Modulation of Secondary Metabolite Production by
; TITLE OF INVENTION: Zinc Binuclear Cluster Proteins
; FILE REFERENCE: 14184-019US1
; CURRENT APPLICATION NUMBER: US/10/149,310
; CURRENT FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: PCT/US01/29288
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 60/233,564
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 147
; LENGTH: 2595
; TYPE: DNA
; ORGANISM: Kluyveromyces marxianus var. lactis
US-10-149-310-147

Query Match 67.7%; Score 21; DB 17; Length 2595;
Best Local Similarity 82.8%; Pred. No. 1.1e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTAATATTATTCAATAATATTGCTCACA 30
DB 1027 CTATACTAAACAATAATATTACTCTAA 1055

RESULT 15
US-10-473-126-189/c
; Sequence 189, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 189
; LENGTH: 5333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-189

Query Match 67.7%; Score 21; DB 18; Length 5333;
Best Local Similarity 82.8%; Pred. No. 1.2e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTATATTATTCAATAATATTGCTCACAA 30
|||||
Db 2983 CTATATTATTAAATAATATTATTCTTAA 2955

Search completed: April 24, 2005, 05:52:17
Job time : 253.62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:31:54 ; Search time 1642.13 Seconds
(without alignments)
718.576 Million cell updates/sec

Title: US-10-039-183A-15

Perfect score: 31
Sequence: 1 gctaattattcaataattgtctcaac 31

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gse1:
9: gb_gse2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	23.6	76.1	958	8	CC266736 CH261-90B
2	23.2	74.8	550	1	AA516958 vh85d10.r
3	23.6	72.9	418	8	BH612485 SALK 0328
4	22.2	71.6	751	9	AG394257 Mus muscu
5	22	71.0	1092	9	AG088415 Pan trogl
c 6	21.8	70.3	589	9	BX207173 Danio rer
c 7	21.6	69.7	416	1	AL381307
c 8	21.6	69.7	434	1	AL381308 MtBC019B1
9	21.6	69.7	524	9	CR325923 Medicago
c 10	21.6	69.7	557	9	TA68E10P
c 11	21.6	69.7	567	9	TA68F10P
12	21.6	69.7	580	8	AQ491708
c 13	21.6	69.7	720	8	BH433207 BOGSM15TF
c 14	21.6	69.7	782	8	BZ503289 BONGE68TR
15	21.6	69.7	832	8	BZ820968 PUFQ046TB
c 16	21.6	69.7	885	8	BZ721736 PUDAH76TB
17	21.6	69.7	948	9	CG110457 FUJCC47TD
c 18	21.6	69.7	990	9	CL010460 ZMMBBb055
19	21.4	69.0	399	9	AL977636 Danio rer
c 20	21.4	69.0	448	8	AZ161299 SP 0069 B
21	21.4	69.0	677	7	CO578276 TvEST091B
22	21.4	69.0	691	7	CF114243 Shultzomi
c 23	21.4	69.0	699	8	AZ138814 SP 0175 A
24	21.4	69.0	816	2	BF065214 HV_Ceb002

25	21.4	69.0	826	2	B8559176	B8559176 HV_Ceb002
c 26	21.2	68.4	223	2	B880225	B880225 CM2-ET012
27	21.2	68.4	489	6	CD812340	CD812340 BN10.020N
28	21.2	68.4	489	6	CD814287	CD814287 BN15.022K
29	21.2	68.4	489	6	CD815821	CD815821 BN15.027J
c 30	21.2	68.4	734	6	CB619111	CB619111 OSIIEa03H
c 31	21.2	68.4	1039	8	CC239056	CC239056 CH261-120
c 32	21.2	68.4	1202	9	AG277970	AG277970 Mus muscu
c 33	21	67.7	231	8	CC027685	CC027685 3591.1_6
34	21	67.7	402	6	CB170034	CB170034 CZ1603001
35	21	67.7	416	4	B1208331	B1208331 EST526371
36	21	67.7	422	4	B1208557	B1208557 EST526597
c 37	21	67.7	437	9	CE024196	CE024196 tigr-gss-
c 38	21	67.7	460	2	BF602438	BF602438 267670 MA
c 39	21	67.7	510	2	BF602439	BF602439 267671 MA
40	21	67.7	553	8	BZ276072	BZ276072 CH230-450
41	21	67.7	649	5	BW048970	BW048970 BW048970
42	21	67.7	666	8	BZ758001	BZ758001 PUFCD85TB
43	21	67.7	671	5	BP131560	BP131560 BP131560
44	21	67.7	678	7	CN514609	CN514609 2464.1 Tu
45	21	67.7	688	9	AG151595	AG151595 Pan trogl

ALIGNMENTS

CC266736 958 bp DNA linear GSS 13-MAY-2003
CH261-90B21_Sp6.1 CH261 Gallus gallus genomic clone CH261-90B21,
genomic survey sequence.
ACCESSION CC266736
VERSION CC266736.1 GI:30616030
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 958)
AUTHORS Kremetzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE Gallus gallus BAC End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 7
High quality sequence stop: 648.
Location/Qualifiers
1. .958
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clones="CH261-90B21"
/sex="female"
/clone_lib="CH261"
/clone_line="UCD001, inbred 256"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chori.org/bacpac"

Query Match 76.1%; Score 23.6; DB 8; Length 958;
Best Local Similarity 86.7%; Pred No. 1.9e+02;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 GCTAATATTATTCAATAATATTGCTCAAA 30

Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@tc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY : pBACe3.6

Vector : EcorI

R.Site 1 : EcorI

R.Site 2 : EcorI

Location/Qualifiers

1..751
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMG01-213C16.TJ"
 /sex="male"
 /tissue type="mixture of kidney and spleen"
 /clone_lib="MSMG01 Mouse Male BAC Library"

FEATURES

source

ORIGIN

Query Match 71.6%; Score 22.2; DB 9; Length 751;
 Best Local Similarity 88.9%; Pred. No. 6e+02;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 TAATATTATTCATTAATATGCTCACA 29

Db 656 TAATATTATTCATTAATTAATCTCTCA 682

RESULT 5

AG088415

LOCUS

DEFINITION Pan troglodytes DNA, clone: PTB-087F20.F, genomic survey sequence.

AG088415

ACCESSION

VERSION AG088415.1 GI:16640217

KEYWORDS

SOURCE

ORGANISM Pan troglodytes (chimpanzee)

DEFINITION Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

2 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 1092)

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

Submitted (02-AUG-2001)

Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:chimbases@tc.riken.go.jp, URL:http://hsp.gsc.riken.go.jp/,

tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1..1092

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="PTB-087F20.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

FEATURES

source

ORIGIN

Query Match 71.0%; Score 22; DB 9; Length 1092;
 Best Local Similarity 83.3%; Pred. No. 7.1e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 CTAATATTATTCATTAATATGCTCACAAC 31

Db 297 CTAATATAATTAATAAATTTCTCACTAC 326

RESULT 6

EX207173/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

DEFINITION

Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: segref@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
 Biologie Moleculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
 Mt-est@toulouse.inra.fr Website :
<http://sequence.toulouse.inra.fr/Mtruncatula.html>).

FEATURES source

```

1. .416
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBC019B1"
/tissue_type="arbuscular mycorrhiza"
/dev_stages="harvested 3 weeks post inoculation with Glomus intraradices"
/clone_lib="MtBC"
/notes="Vector: pBluescript pSK; Site 1: EcoRI; Site 2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Epiisses soil : 2/3 calcined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LPAB). The plants were watered every day and twice a week with a modified nutrient Long Ashton solution without phosphate but with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France). Note : EST may be of fungal origin."

```

ORIGIN

```

Query Match      69.7%; Score 21.6; DB 1; Length 416;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAAATATTTCATAATAATATGCTCAAA 30
    ||| ||||| ||||| ||| |||||
Db 322 TAAATATTTCATAATAATATGCTCAAA 295

```

```

RESULT 8
AL381308/c
LOCUS      AL381308      434 bp      mRNA      linear      EST 03-AUG-2000
DEFINITION Medicago truncatula cDNA clone MtBC019B1 T7, mRNA
sequence.
ACCESSION  AL381308.1      GI:9681059
VERSION     AL381308
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
ORGANISM    Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1 (bases 1 to 434)
AUTHORS    Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D.,
            Gianinazzi-Pearson,V. and Gamas,P.
            Medicago truncatula ESTs from endomycorrhizal roots
            Unpublished (2000)
CONTACT    Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

```

Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
 Biologie Moleculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
 Mt-est@toulouse.inra.fr Website :
<http://sequence.toulouse.inra.fr/Mtruncatula.html>).

FEATURES source

```

1. .434
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBC019B1"
/tissue_type="arbuscular mycorrhiza"
/dev_stages="harvested 3 weeks post inoculation with Glomus intraradices"
/clone_lib="MtBC"
/notes="Vector: pBluescript pSK; Site 1: EcoRI; Site 2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Epiisses soil : 2/3 calcined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LPAB). The plants were watered every day and twice a week with a modified nutrient Long Ashton solution without phosphate but with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France). Note : EST may be of fungal origin."

```

ORIGIN

```

Query Match      69.7%; Score 21.6; DB 1; Length 434;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAAATATTTCATAATAATATGCTCAAA 30
    ||| ||||| ||||| ||| |||||
Db 334 TAAATATTTCATAATAATATGCTCAAA 307

```

```

RESULT 9
CR325923
LOCUS      CR325923      524 bp      DNA      linear      GSS 01-MAR-2004
DEFINITION Medicago truncatula BAC ends cultivar Jemalong AL7 of Medicago
truncatula, genomic survey sequence.
ACCESSION  CR325923
VERSION     CR325923.1      GI:44872067
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
ORGANISM    Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1 (bases 1 to 524)
AUTHORS    Genoscope.
            Direct Submission
            Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr
            Location/Qualifiers
FEATURES source
1. .524
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="Jemalong AL7"
/db_xref="taxon:3880"
/clone_lib="MTE1"

```

/note="Vector: pindigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
; Debelle F. and Chalhoub B.-Genoscope sequence ID :
mtei-53L24RM1"

ORIGIN

Query Match 69.7%; Score 21.6; DB 9; Length 524;
Best Local Similarity 85.7%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30

||||| ||||||| ||||||| ||||||| |||||||

Db 218 TAATAATCTTCAATAATATTGCTCAAA 245

RESULT 10

TA68E10P/c

LOCUS

DEFINITION T. brucei sheared genomic DNA clone 68e10, forward sequence,
genomic survey sequence.

ACCESSION

AL457511

VERSION

AL457511.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Trypanosoma brucei

Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE

AUTHORS

Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,

Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,

Melville, S.E., Rajandream, M.A. and Barrell, B.G.

Direct Submission

Submitted (10-DEC-2000)

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

nh@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared

to give a tight size distribution (

4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at http://www.sanger.ac.uk/Projects/T_brucei/.

Location/Qualifiers

1..557

/organism="Trypanosoma brucei"

/mol_type="genomic DNA"

/strain="TREU927"

/db_xref="taxon:5691"

/clones="68e10"

ORIGIN

Query Match 69.7%; Score 21.6; DB 9; Length 557;

Best Local Similarity 85.7%; Pred. No. 9.9e+02;

Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30

||||| ||||||| ||||||| ||||||| |||||||

Db 434 TAATATTATTGATATTGCTCAAA 407

RESULT 11

TA68F10P/c

LOCUS

DEFINITION T. brucei sheared genomic DNA clone 68f10, forward sequence,

genomic survey sequence.

ACCESSION

AL457517

VERSION

AL457517.1

KEYWORDS

GSS.

SOURCE
ORGANISM

Trypanosoma brucei

Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE

AUTHORS

Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,

Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,

Melville, S.E., Rajandream, M.A. and Barrell, B.G.

Direct Submission

Submitted (10-DEC-2000)

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

nh@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared

to give a tight size distribution (

4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at http://www.sanger.ac.uk/Projects/T_brucei/.

Location/Qualifiers

1..567

/organism="Trypanosoma brucei"

/mol_type="genomic DNA"

/strain="TREU927"

/db_xref="taxon:5691"

/clones="68f10"

ORIGIN

Query Match 69.7%; Score 21.6; DB 9; Length 567;

Best Local Similarity 85.7%; Pred. No. 9.9e+02;

Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCAATAATATTGCTCAAA 30

||||| ||||||| ||||||| ||||||| |||||||

Db 434 TAATATTATTGATATTGCTCAAA 407

RESULT 12

TA68F10P

LOCUS

DEFINITION RPCI-11-271F14.TJ RPCI-11 Homo sapiens genomic clone

RPCI-11-271F14, genomic survey sequence.

ACCESSION

AQ491708

VERSION

AQ491708.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 580)

Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and

Venter, J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Map Building

Unpublished (1997)

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.

Seq primer: SP6
Class: BAC ends.

FEATURES

source
Location/Qualifiers
1..580
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7603813"
/db_xref="taxon:9606"
/clone="RPC1-11-271F14"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPC1-11"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
RPC111 Human Male BAC Library"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 580;
Best Local Similarity 85.7%; Pred. No. 9.e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATAATATTCGTCAAA 30
|||||
Db 284 TAATATTATTCATAATATTCGTCAAA 311

RESULT 13

BH433207/c
LOCUS
DEFINITION BH433207 720 bp DNA linear GSS 12-DEC-2001
BOGSW15TF BOGS Brassica oleracea genomic clone BOGSW15, genomic
survey sequence.
ACCESSION BH433207
VERSION BH433207.1 GI:17618928
KEYWORDS
SOURCE
ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 720)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOGSW15TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.

FEATURES

source
Location/Qualifiers
1..720
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGSW15"
/clone_lib="BOGS"
/note="Vector: pBACe3.6; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pBACe3 using BstXI linkers"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 720;
Best Local Similarity 85.7%; Pred. No. 1.e+03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATAATATTCGTCAAA 30
|||||
Db 318 TAATATTCTCCATTATAGTCACAA 291

RESULT 14

BZ503289/c
LOCUS
DEFINITION BONGE68TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONGE68,
genomic survey sequence.
ACCESSION BZ503289
VERSION BZ503289.1 GI:27021620
KEYWORDS
SOURCE
ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 782)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BONGE68TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES

source
Location/Qualifiers
1..782
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONGE68"
/clone_lib="BO_1.6_2_KB_tot"
/note="Vector: pBACe3.6; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pBACe3 using BstXI linkers"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 782;
Best Local Similarity 85.7%; Pred. No. 1.e+03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TAATATTATTCATAATATTCGTCAAA 30
|||||
Db 416 TAATATTCTCCATTATAGTCACAA 389

RESULT 15

BZ820968
LOCUS
DEFINITION PUFGO46TB ZM_0.6_1.0_KB_Zea mays genomic clone ZMMBT314H19,
genomic survey sequence.
ACCESSION BZ820968
VERSION BZ820968.1 GI:29035790
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 832)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUFGO46TD
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843

Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TR
 Class: sheared ends.

FEATURES

Location/Qualifiers

1..832
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBTa314H19"
 /clone_lib="ZM 0.6-1.0 KB"
 /note="Vector: PCR-TOPO, Site_1: EcoRI; 0.6-1.0 kb high
 Cor selected genomic DNA library"

ORIGIN

Query Match 69.7%; Score 21.6; DB 8; Length 832;
 Best Local Similarity 85.7%; Pred. No. 1e+03;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 4 AATATTATTCAATAATATTGCTCAAC 31
 ||||| ||||| ||||| |||||
 Db 770 AATATTATGAATAATATAGCTAACAC 797

Search completed: April 24, 2005, 05:28:22
 Job time : 1646.13 secs

THIS PAGE BLANK (U9PT6)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 10:03:10 ; Search time 177.401 Seconds
(without alignments)
900.968 Million cell updates/sec

Title: US-10-039-183A-16
Perfect score: 27
Sequence: 1 99ggaatacaaatgcaaaaagaaag 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	27	AAV07971	Helicobac
2	27	100.0	1448	AAV07964	AAV07964 Helicobac
C 3	21.8	80.7	12834	AAK73037	AAK73037 Human imm
C 4	21.8	80.7	256190	13 ABD33276	Abd33276 Human can
5	21.4	79.3	82660	11 ACN45192	Acn45192 Mouse gen
6	21.2	78.5	1200	9 ADB06507	ADB06507 Alloiococ
C 7	21.2	78.5	110000	9 ADB12064_00	ADB12064 Alloiococ
C 8	21.2	78.5	110000	9 ADB12064_01	Continuation (2 of
C 9	20.8	77.0	2004	4 AAL37651	Aal37651 Human mus
C 10	20.8	77.0	2004	4 AAL37650	Aal37650 Human mus
C 11	20.8	77.0	2004	8 ABX60638	ABX60638 cDNA enco
C 12	20.8	77.0	2004	8 ABX60639	ABX60639 cDNA enco
C 13	20.8	77.0	2004	12 ADJ31388	Adj31388 Human mus
C 14	20.8	77.0	2004	12 ADJ31389	Adj31389 Human mus
15	20.8	77.0	110000	8 ADD53224_1	Continuation (2 of
16	20.6	76.3	10999	4 ABL03656	Abi03656 Drosophil
C 17	20.6	76.3	175590	10 ADD50650	Add50650 BAC sequ
18	20.4	75.6	5181	4 ABL10975	Abi10975 Drosophil
C 19	20.4	75.6	14974	4 ABL10974	Abi10974 Drosophil
C 20	20.2	74.8	893	3 AAC45523	Aac45523 Arabidops

C 21	20.2	74.8	897	3 AAC37376	Aac37376 Arabidops
22	20.2	74.8	7492	4 AAC90510	Aac90510 Mouse fac
23	20.2	74.8	7493	2 AAV12115	Aav12115 Mus muscu
24	20.2	74.8	7493	2 AAV25812	Aav25812 Murine fa
25	20.2	74.8	7493	2 AAX91164	Aax91164 Mouse fac
26	20.2	74.8	7493	10 ADK48891	Adk48891 Wild type
C 27	20.2	74.8	16092	4 AAK73420	Aak73420 Human imm
28	20.2	74.8	32183	4 AAI19267	Aai19267 Human exc
29	20.2	74.8	32183	5 AAI63617	Aai63617 Human kid
30	19.8	73.3	2988	8 ADA71057	Ada71057 Rice gene
C 31	19.8	73.3	3008	10 ADF60626	Adf60626 Melon gal
32	19.8	73.3	163701	13 ABD33351	Abd33351 Murine ca
33	19.6	72.6	359	4 AAI91071	Aai91071 Human pol
34	19.6	72.6	529	13 ACN56029	Acn56029 Cotton an
35	19.6	72.6	954	13 ADS59095	AdS59095 Bacterial
36	19.6	72.6	1140	10 ADH84737	Adh84737 Enterococ
37	19.6	72.6	1187	8 ACC49506	Acc49506 Tumour-as
38	19.6	72.6	2148	4 AAF57111	Aaf57111 A. thalia
39	19.6	72.6	2547	4 AAF58469	Aaf58469 AtCNGC1/D
C 40	19.6	72.6	4326	13 ACN42452	Acn42452 Human dia
41	19.6	72.6	5996	4 AAF58468	Aaf58468 AtCNGC1/D
C 42	19.6	72.6	21591	2 AAX13047	Aax13047 Enterococ
C 43	19.6	72.6	21591	6 ABS98842	AbS98842 Enterococ
C 44	19.6	72.6	110000	13 ABD32629_3	Continuation (4 of
C 45	19.6	72.6	240000	8 ACD13446	ACD13446 Human DNA

ALIGNMENTS

RESULT 1

AAV07971

ID AAV07971 standard; DNA; 27 BP.

XX AC AAV07971;

XX DT 25-MAR-2003 (revised)

XX DT 02-FEB-1999 (first entry)

XX DE Helicobacter pylori polypeptide GHPO 750 5' DNA primer.

XX KW GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;

XX KW primer; ss.

XX OS Synthetic.

XX OS Helicobacter pylori.

XX PN WO9843479-A1.

XX PD 08-OCT-1998.

XX PF 31-MAR-1998; 98WO-US006421.

XX PR 01-APR-1997; 97US-00831310.

XX PR 01-APR-1997; 97US-00834666.

XX PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WI WI; 1998-568251/48.

XX New isolated Helicobacter polynucleotides - used to develop products for

XX the diagnosis, prevention and treatment of Helicobacter infections and

XX gastroduodenal diseases.

XX Claim 5; Page 156; 184pp; English.

XX This 5' primer is used with a 3' primer (see AAV07972) in the PCR
XX amplification of Helicobacter, e.g. Helicobacter pylori, genomic DNA in
XX order to obtain DNA (see AAV07964) encoding the a 50 kDa polypeptide (see
XX AAW73035) designated GHPO 750. The isolated polynucleotide, and encoded

CC polypeptide, can be used to develop vaccines for the treatment and
CC prevention of Helicobacter infections. (Updated on 25-MAR-2003 to correct
CC PI field.)

XX Sequence 27 BP; 16 A; 2 C; 7 G; 2 T; 0 U; 0 Other;
Query Match 100.0%; Score 27; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGAGAAATACAAATGGCAAAAGAAAG 27
|||||
Db 1 GGAGAAATACAAATGGCAAAAGAAAG 27
|||||

RESULT 2

AAV07964
ID AAV07964 standard; DNA; 1448 BP.

XX AC AAV07964;

XX 25-MAR-2003 (revised)
DT 02-FEB-1999 (first entry)

XX Helicobacter pylori 50 kDa polypeptide GHPO 750 DNA.

XX GHPO 750; infection; gastritis; ulcer; vaccine; diagnosis; therapy; ss.

XX Helicobacter pylori.

XX Key Location/Qualifiers
FT CDS 118..1317
FT /*tag= a

XX WO9843479-A1.

XX 08-OCT-1998.

XX 31-MAR-1998; 98WO-US006421.

XX 01-APR-1997; 97US-00831310.

XX 01-APR-1997; 97US-00834666.

XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;

XX WPI; 1998-568251/48.

XX P-PSDB; AAV73035.

XX New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastroduodenal diseases.

XX Claim 1; Page 150-152; 184pp; English.

XX This DNA sequence codes for a 50 kDa Helicobacter pylori polypeptide (see
CC AAV73035) designated GHPO 750. A polynucleotide encoding GHPO 750 can be
CC obtained from genomic DNA by PCR amplification (see AAV07971-72. The
CC invention provides polynucleotides (see AAV72001, AAV07912-21 and
CC AAV07963-64) encoding a family 76 kDa Helicobacter polypeptides (see
CC AAV73022-32), GHPO 750 and a 32 kDa polypeptide (see AAV73034). These
CC polynucleotides were initially identified in a search of H. pylori
CC genomic databases. DNA cassettes for expression of the Helicobacter
CC proteins (unprocessed or mature forms) in prokaryotic or eukaryotic cells
CC are provided. The polynucleotides can be used in vaccines to prevent or
CC treat Hb infection in a mammal. Viral (especially poxvirus) or bacterial
CC vectors are used. Products and methods of the invention allow treatment
CC and prevention of gastroduodenal diseases associated with Hb infections,
CC including acute, chronic, and atrophic gastritis, and peptic ulcer
CC diseases, e.g. gastric and duodenal ulcers. Diagnostic and detection
CC methods are also provided. GHPO 750 was demonstrated to be a protective

CC antigen. (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 1448 BP; 466 A; 237 C; 364 G; 381 T; 0 U; 0 Other;

Query Match 100.0%; Score 27; DB 2; Length 1448;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGAGAAATACAAATGGCAAAAGAAAG 27
|||||
Db 106 GGAGAAATACAAATGGCAAAAGAAAG 132
|||||

RESULT 3

AAK73037/C

ID AAK73037 standard; DNA; 12834 BP.

XX AC AAK73037;

XX 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27849.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-020515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-021513P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 18-AUG-2000; 2000US-0226279P.

XX 22-AUG-2000; 2000US-0226681P.

XX 22-AUG-2000; 2000US-0226686P.

XX 23-AUG-2000; 2000US-0227182P.

XX 30-AUG-2000; 2000US-0227009P.

XX 01-SEP-2000; 2000US-0228944P.

XX 01-SEP-2000; 2000US-0229287P.

XX 01-SEP-2000; 2000US-0229343P.

XX 01-SEP-2000; 2000US-0229344P.

```
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229503P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 14-SEP-2000; 2000US-0233066P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249220P.
PR 17-NOV-2000; 2000US-0249221P.
PR 17-NOV-2000; 2000US-0249222P.
PR 17-NOV-2000; 2000US-0249223P.
PR 17-NOV-2000; 2000US-0249224P.
PR 17-NOV-2000; 2000US-0249225P.
PR 17-NOV-2000; 2000US-0249226P.
PR 17-NOV-2000; 2000US-0249227P.
PR 17-NOV-2000; 2000US-0249228P.
PR 17-NOV-2000; 2000US-0249229P.
PR 17-NOV-2000; 2000US-0249230P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251859P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 27849; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to prevent,
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 12834 BP; 3916 A; 2492 C; 2530 G; 3896 T; 0 U; 0 Other;
XX Query Match 80.7%; Score 21.8; DB 4; Length 12834;
XX Best Local Similarity 92.0%; Pred. No. 2e+02;
XX Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GGAGAAATACAAATGGCAAAAGAAA 25
Db 6047 GGAGAAATACAAATGGCAAAAGAAA 6023
RESULT 4
ABD33276/c
ID ABD33276 standard; DNA; 256190 BP.
XX
XX AC ABD33276;
XX
XX DT 18-NOV-2004 (first entry)
XX
```

DE Human cancer-associated (CA) gene HD07-048.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
XX Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
XX 17-DEC-2002; 2002US-00322281.
PR
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX
XX WPI; 2004-499109/47.
XX
XX Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
XX Claim 16; SEQ ID NO 320; 182pp; English.
XX
XX The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 256190 BP; 75634 A; 45403 C; 48850 G; 84855 T; 0 U; 1448 Other;
SQ
Query Match 80.7%; Score 21.8; DB 13; Length 256190;
Best Local Similarity 92.0%; Pred. NO. 2.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGAGAAATACAAATGGCAAAAGAAA 25
DB 243196 GGAGAAATACAAATGGCAAAAGAAA 243172
RESULT 5
ACN45192
ID ACN45192 standard; DNA; 82660 BP.
XX
XX ACN45192;
XX
XX 18-NOV-2004 (first entry)
DT
XX Mouse genomic sequence MCG9397.
DE
XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
KW
XX Mus musculus.
OS
XX WO2003073826-A2.
PN
XX 12-SEP-2003.
PD
XX 28-FEB-2003; 2003WO-US006235.
PF
Human cancer-associated (CA) gene HD07-048.
XX
XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX ds; cancer; cytostatic.
XX
XX Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
XX 17-DEC-2002; 2002US-00322281.
PR
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW, Malandro MS;
XX
XX WPI; 2004-499109/47.
XX
XX Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
XX Claim 16; SEQ ID NO 320; 182pp; English.
XX
XX The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 256190 BP; 75634 A; 45403 C; 48850 G; 84855 T; 0 U; 1448 Other;
SQ
Query Match 80.7%; Score 21.8; DB 13; Length 256190;
Best Local Similarity 92.0%; Pred. NO. 2.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGAGAAATACAAATGGCAAAAGAAA 25
DB 243196 GGAGAAATACAAATGGCAAAAGAAA 243172
RESULT 5
ACN45192
ID ACN45192 standard; DNA; 82660 BP.
XX
XX ACN45192;
XX
XX 18-NOV-2004 (first entry)
DT
XX Mouse genomic sequence MCG9397.
DE
XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
KW
XX Mus musculus.
OS
XX WO2003073826-A2.
PN
XX 12-SEP-2003.
PD
XX 28-FEB-2003; 2003WO-US006235.
PF
Human cancer-associated (CA) gene HD07-048.
XX
XX Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX ds; cancer; cytostatic.
XX
XX Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
XX 17-DEC-2002; 2002US-00322281.
PR
XX (SAGR-) SAGRES DISCOVERY INC.
XX
XX Morris DW;
XX
XX WPI; 2003-328604/31.
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
XX Claim 1; SEQ ID NO 2017; Opp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US200218586A1, for which no sequence data was published
XX
XX Sequence 82660 BP; 19374 A; 15352 C; 16570 G; 20443 T; 0 U; 10921 Other;
SQ
Query Match 79.3%; Score 21.4; DB 11; Length 82660;
Best Local Similarity 95.7%; Pred. NO. 3.1e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 AAATACAAATGGCAAAAGAAAAG 27
DB 31339 ACATACAAATGGCAAAAGAAAAG 31361
RESULT 6
ADB06507
ID ADB06507 standard; DNA; 1200 BP.
XX
XX ADB06507;
AC
XX 20-NOV-2003 (first entry)
DT
XX Alloicoccus otitis antigenic protein encoding DNA SEQ ID NO:447.
DE
XX Alloicoccus otitis;
KW Alloicoccus otitis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive Bacterium; infection; gene; ds.
XX
XX Alloicoccus otitis.
OS
XX WO2003048304-A2.
PN
XX 12-JUN-2003.
PD
XX 25-NOV-2002; 2002WO-US036123.
PF
XX 29-NOV-2001; 2001US-0333777P.
PR
XX 18-NOV-2002; 2002US-0426742P.
XX
XX (AMHP) WYETH HOLDINGS CORP.
PA
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
PI
XX WPI; 2003-505284/47.
XX
XX P-FSDB; ADB06508.
DR
XX New Alloicoccus otitis polynucleotides and polypeptides, useful for
XX treating and diagnosing diseases, drug screening assays and monitoring of
XX

PT effects during drug clinical trials.

XX Claim 7; SEQ ID NO 447; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of
 CC Alloiococcus otitidis genomic DNA, which encodes an antigenic protein.
 CC Alloiococcus otitidis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
 CC expression vector comprising the novel isolated polynucleotide (I), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against Alloiococcus otitidis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying Alloiococcus
 CC otitidis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (I) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting Alloiococcus
 CC otitidis. The present sequence encodes an Alloiococcus otitidis antigen
 CC protein from the present invention.

XX SQ Sequence 1200 BP; 343 A; 260 C; 264 G; 333 T; 0 U; 0 Other;

Query Match 78.5%; Score 21.2; DB 9; Length 1200;

Best Local Similarity 88.5%; Pred. No. 3e+02; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGAAATACAAATGGCAAAAGAAAA 26

Db 4 GGAAAAATAACATGGCAAAAGAAAA 29

RESULT 7

ADBI2064_00/c

WP Sequence split into 18 fragments LOCUS ADBI2064 Accession Adbi2064

Fragment Name	Begin	End
ADBI2064_00	1	110000
ADBI2064_01	100001	210000
ADBI2064_02	200001	310000
ADBI2064_03	300001	410000
ADBI2064_04	400001	510000
ADBI2064_05	500001	610000
ADBI2064_06	600001	710000
ADBI2064_07	700001	810000
ADBI2064_08	800001	910000
ADBI2064_09	900001	1010000
ADBI2064_10	1000001	1110000
ADBI2064_11	1100001	1210000
ADBI2064_12	1200001	1310000
ADBI2064_13	1300001	1410000
ADBI2064_14	1400001	1510000
ADBI2064_15	1500001	1610000
ADBI2064_16	1600001	1710000
ADBI2064_17	1700001	1754382

ID ADBI2064 standard; DNA; 1754382 BP.

XX ADBI2064;

AC 20-NOV-2003 (first entry)

XX Alloiococcus otitis entire genome sequence SEQ ID NO:6651.

DE Alloiococcus otitidis; antigenic protein; immunogenic; immunisation;

XX

KW gene therapy; Gram-positive bacterium; infection; gene; ds.

XX OS Alloiococcus otitis.

XX WO2003048304-A2.

XX 12-JUN-2003.

XX 25-NOV-2002; 2002WO-US036123.

XX 29-NOV-2001; 2001US-0333777P.

XX 18-NOV-2002; 2002US-0426742P.

XX (AMHP) WYETH HOLDINGS CORP.

XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

XX New Alloiococcus otitidis polynucleotides and polypeptides, useful for
 PT treating and diagnosing diseases, drug screening assays and monitoring of
 PT effects during drug clinical trials.

XX Example 3; SEQ ID NO 6651; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of
 CC Alloiococcus otitidis genomic DNA, which encodes an antigenic protein.
 CC Alloiococcus otitidis is a Gram-positive bacterium. Also described: (1)
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
 CC expression vector comprising the novel isolated polynucleotide (I), its
 CC complement, degenerate variant or fragment; (3) a genetically engineered
 CC host cell, transfected, transformed or infected with the vector of (2);
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
 CC composition comprising the polypeptide, its complement, biological
 CC equivalent or fragment, or the polynucleotide that is comprised in the
 CC expression vector; (6) a pharmaceutical composition comprising the
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)
 CC immunising against Alloiococcus otitidis by administering to a host the
 CC immunogenic composition; (9) detecting and/or identifying Alloiococcus
 CC otitidis in the biological sample; (10) a kit comprising a container
 CC containing the novel polynucleotide, its degenerate variant or fragment,
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the
 CC genetically engineered host cell under conditions suitable to produce the
 CC polypeptide from the culture. (I) can be used in gene therapy. The
 CC polynucleotides, polypeptides, antibodies and compositions of the present
 CC invention can be used for treating and diagnosing diseases, drug
 CC screening assays and monitoring of effects during drug clinical trials.
 CC The polynucleotides are useful for expressing and detecting Alloiococcus
 CC otitidis. The present sequence represents the entire genome of
 CC Alloiococcus otitidis, which is given in the exemplification of the
 CC present invention.

XX SQ Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 0Other;

Query Match 78.5%; Score 21.2; DB 9; Length 110000;

Best Local Similarity 88.5%; Pred. No. 3.7e+02;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGAAATACAAATGGCAAAAGAAAA 26

Db 104256 GGAAAAATAACATGGCAAAAGAAAA 104231

RESULT 8

ADBI2064_01/c

Continuation (2 of 18) of ADBI2064 from base 100001 (Alloiococcus otitis entire genome B

WP Sequence split into 18 fragments LOCUS ADBI2064 Accession Adbi2064

Fragment Name	Begin	End
ADBI2064_00	1	110000
ADBI2064_01	100001	210000
ADBI2064_02	200001	310000
ADBI2064_03	300001	410000

WP	ADBI2064_04	400001	510000	PR	14-AUG-2000;	2000US-0225268P.
WP	ADBI2064_05	500001	610000	PR	14-AUG-2000;	2000US-0225270P.
WP	ADBI2064_06	600001	710000	PR	14-AUG-2000;	2000US-0225447P.
WP	ADBI2064_07	700001	810000	PR	14-AUG-2000;	2000US-0225757P.
WP	ADBI2064_08	800001	910000	PR	14-AUG-2000;	2000US-0225758P.
WP	ADBI2064_09	900001	1010000	PR	14-AUG-2000;	2000US-0225759P.
WP	ADBI2064_10	1000001	1110000	PR	18-AUG-2000;	2000US-0226279P.
WP	ADBI2064_11	1100001	1210000	PR	22-AUG-2000;	2000US-0226691P.
WP	ADBI2064_12	1200001	1310000	PR	22-AUG-2000;	2000US-0226868P.
WP	ADBI2064_13	1300001	1410000	PR	22-AUG-2000;	2000US-0227182P.
WP	ADBI2064_14	1400001	1510000	PR	23-AUG-2000;	2000US-0227009P.
WP	ADBI2064_15	1500001	1610000	PR	30-AUG-2000;	2000US-0228924P.
WP	ADBI2064_16	1600001	1710000	PR	01-SEP-2000;	2000US-0229287P.
WP	ADBI2064_17	1700001	1754382	PR	01-SEP-2000;	2000US-0229343P.
Query Match						2000US-0229344P.
Best Local Similarity						2000US-0229345P.
Matches 23; Conservative						2000US-0229509P.
88.5%; Score 21.2; DB 9; Length 110000;						2000US-0229513P.
Pred. No. 3.7e+02;						2000US-0230437P.
Indels 0; Gaps 0;						2000US-0230438P.
Mismatches 3;						2000US-0231242P.
Qy	1	CGAGAAATACAAATGCGCAAGAAAA	26	PR	06-SEP-2000;	2000US-0231243P.
						2000US-0231244P.
4256						2000US-0231413P.
GGAATAAATACATGCGCAAGAAAA						2000US-0231414P.
4231						2000US-0231988P.
RESULT 9						2000US-0232397P.
AAL37651/c						2000US-0232398P.
ID	AAL37651 standard; DNA; 2004 BP.					2000US-0232399P.
XX	AC					2000US-0232400P.
XX	AAL37651;					2000US-0232401P.
XX	08-JAN-2002 (first entry)					2000US-0233063P.
XX	Human musculoskeletal system related polynucleotide SEQ ID NO 4016.					2000US-0233064P.
XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;					2000US-0233065P.
XX	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;					2000US-0234274P.
XX	vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;					2000US-0234997P.
XX	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;					2000US-0235484P.
XX	neurological disease; infection; human; secreted protein;					2000US-0235834P.
XX	musculoskeletal system; ds.					2000US-0236327P.
OS	Homo sapiens.					2000US-0236328P.
XX	WO200155367-A1.					2000US-0236368P.
XX	02-AUG-2001.					2000US-0236369P.
XX	17-JAN-2001; 2001WO-US001338.					2000US-0237037P.
XX	31-JAN-2000; 2000US-0179065P.					2000US-0237038P.
XX	04-FEB-2000; 2000US-0180628P.					2000US-0237039P.
XX	24-FEB-2000; 2000US-0184664P.					2000US-0237040P.
XX	02-MAR-2000; 2000US-0186350P.					2000US-0239935P.
XX	16-MAR-2000; 2000US-0189874P.					2000US-0239937P.
XX	17-MAR-2000; 2000US-0190076P.					2000US-0240960P.
XX	18-APR-2000; 2000US-0198123P.					2000US-0241221P.
XX	19-MAY-2000; 2000US-0205515P.					2000US-0241785P.
XX	07-JUN-2000; 2000US-0209467P.					2000US-0241786P.
XX	28-JUN-2000; 2000US-0214886P.					2000US-0241878P.
XX	30-JUN-2000; 2000US-0215135P.					2000US-0241808P.
XX	07-JUL-2000; 2000US-0216647P.					2000US-0241809P.
XX	07-JUL-2000; 2000US-0216880P.					2000US-0241826P.
XX	11-JUL-2000; 2000US-0217496P.					2000US-0244617P.
XX	11-JUL-2000; 2000US-0218290P.					2000US-0246474P.
XX	26-JUL-2000; 2000US-0220963P.					2000US-0246475P.
XX	26-JUL-2000; 2000US-0220964P.					2000US-0246476P.
XX	14-AUG-2000; 2000US-0224518P.					2000US-0246477P.
XX	14-AUG-2000; 2000US-0224519P.					2000US-0246523P.
XX	14-AUG-2000; 2000US-0225213P.					2000US-0246524P.
XX	14-AUG-2000; 2000US-0225214P.					2000US-0246525P.
XX	14-AUG-2000; 2000US-0225266P.					2000US-0246526P.
XX	14-AUG-2000; 2000US-0225267P.					

```
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451937/48.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
XX disorders related to the musculoskeletal system including musculoskeletal
XX cancers and also for testing and detection e.g. diagnosis.
XX
XX Example 2; SEQ ID NO 4016; 781pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
XX (AB03087-AB04109) associated with the musculoskeletal system useful for
XX preventing, treating or ameliorating medical conditions e.g. by protein
XX or gene therapy. The genes are isolated from a range of human tissues
XX disclosed in the specification. The nucleic acids, proteins, antibodies
XX and (ant)agonists are useful in the diagnosis, treatment and prevention
XX of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
XX adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
XX lung, or urogenital; (b) immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; (c) cardiovascular disorders such as
XX myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
XX cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
XX bacterial, fungal and parasitic infections. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2004 BP; 632 A; 428 C; 351 G; 593 T; 0 U; 0 Other;
```

```
Query Match 77.0%; Score 20.8; DB 4; Length 2004;
Best Local Similarity 91.7%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GGAGAAATACAAATGGCAAAAGAA 24
Db 855 GGAGAAATACAAAGGACAAAGAA 832
RESULT 10
AAL37650/c
ID AAL37650 standard; DNA; 2004 BP.
XX
XX AAL37650;
AC
XX
DT 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 4015.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antifungal;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein;
XX musculoskeletal system; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200155367-A1.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001338.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216890P.
XX 11-JUL-2000; 2000US-0217487P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226686P.
XX 23-AUG-2000; 2000US-0227182P.
XX 30-AUG-2000; 2000US-0227009P.
XX 01-SEP-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
```

```
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451937/48.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
XX disorders related to the musculoskeletal system including musculoskeletal
XX cancers and also for testing and detection e.g. diagnosis.
XX
XX Example 2; SEQ ID NO 4015; 781bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
XX preventing, treating or ameliorating medical conditions e.g. by protein
XX or gene therapy. The genes are isolated from a range of human tissues
XX disclosed in the specification. The nucleic acids, proteins, antibodies
XX and (ant)agonists are useful in the diagnosis, treatment and prevention
XX of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
XX adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
XX lung, or urogenital; (b) immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; (c) cardiovascular disorders such as
XX myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
XX cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
XX bacterial, fungal and parasitic infections. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2004 BP; 629 A; 429 C; 353 G; 593 T; 0 U; 0 Other;
XX
XX Query Match 77.0%; Score 20.8; DB 4; Length 2004;
XX Best Local Similarity 91.7%; Pred. No. 4.3e+02;
XX Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GGAGAAATACAAATGGCAAGAA 24
Db 855 GGAGAAATACAAAGGACAAAGAA 832
RESULT 11
ABX60638/C
ID ABX60638 standard; cDNA; 2004 BP.
XX
XX AC ABX60638;
XX
```


DT 26-FEB-2003 (first entry)

XX cDNA encoding novel human musculoskeletal system antigen #2982.

DE Gene; ss; musculoskeletal system antigen; cancer; metastasis;

XX re-vascularisation; thrombosis; arteriosclerosis; mineral content;

XX cardiovascular condition; wound; injury; burn; angiodysplasia; ulcer;

KW post-operative tissue repair; limb regeneration; neuronal growth;

KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

KW AIDS-related complex; chondrocyte growth; bone regeneration;

KW periodontal regeneration; tissue transport; bone graft; skin aging;

KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;

KW cell growth; organ transplant; cell differentiation; body height; weight;

KW hair colour; eye colour; skin; percentage of adipose tissue;

KW pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;

KW depression; tendency for violence; pain; reproductive capability;

KW hormone level; endocrine level; appetite; libido; memory; stress;

KW storage capability; fat content; lipid content; protein content;

KW carbohydrate content; vitamin content; cofactor content;

KW nutritional component.

OS Homo sapiens.

XX US2002147140-A1.

PN 10-OCT-2002.

PD 17-JAN-2001; 2001US-00764877.

PF 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 22-AUG-2000; 2000US-0226868P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 08-SEP-2000; 2000US-0231413P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234997P.

PR 27-SEP-2000; 2000US-0235834P.

PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.

PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239933P.

PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241809P.

PR 01-NOV-2000; 2000US-0244617P.

PR 17-NOV-2000; 2000US-0249299P.

PR 08-DEC-2000; 2000US-0251856P.

PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.

PA (RUBE/) RUBEN S M.

PA (BARA/) BARASH S C.

XX Rosen CA, Ruben SM, Barash SC;

XX WPI; 2003-128199/12.

XX Isolated nucleic acid molecules encoding musculoskeletal system

PT associated polypeptides, useful for detecting disorders, e.g. cancer.

XX Disclosure; SEQ ID NO 4015; 31pp; English.

XX The invention describes an isolated nucleic acid molecule comprising a

CC sequence encoding musculoskeletal system associated polypeptides useful

CC for detecting disorders, e.g., cancer or cancer metastases, in animals or

CC humans. The nucleic acid: stimulates re-vascularisation of ischemic

CC tissues associated with conditions such as thrombosis, arteriosclerosis,

CC and other cardiovascular conditions; treats wounds due to injuries,

CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis

CC and limb regeneration; stimulates neuronal growth; can treat and prevent

CC neuronal damage occurring in certain disorders or neurodegenerative

CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-

CC related complex; stimulates chondrocyte growth, thus they can be used to

CC enhance bone and periodontal regeneration and aid in tissue transports or

CC bone grafts; prevents skin aging due to sunburn by stimulating

CC keratinocyte growth; prevents hair loss, since FGF family members

CC activate hair-forming cells and promotes melanocyte growth; stimulates

CC growth and differentiation of hematopoietic cells and bone marrow cells

CC when used in combination with other cytokines; maintains organs before

CC transplantation or for supporting cell culture of primary tissues;

CC induces tissue of mesodermal origin to differentiate in early embryos;

CC increases or decreases the differentiation or proliferation of embryonic

CC stem cells, besides haematopoietic lineage; modulates mammalian

CC characteristics, such as, body height, weight, hair colour, eye colour,

CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,

CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal

CC state or physical state by influencing biorhythms, cardiac rhythms,

CC depression, tendency for violence, tolerance for pain, reproductive

CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or

CC stress; increases or decreases storage capabilities, fat content, lipid,

CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional

CC components. This sequence encodes a novel human musculoskeletal system

CC antigen. Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from the US patent office at

CC ftp:seqdata.uspto.gov/sequence.html?DocID=20020147140

XX SQ Sequence 2004 BP; 629 A; 429 C; 353 G; 593 T; 0 U; 0 Other;

Query Match 77.0%; Score 20.8; DB 8; Length 2004;

Best Local Similarity 91.7%; Pred. No. 4.3e+02;

Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGGCAAAAGAA 24

Db 855 GGAGAAATACAAAGGACAAAGAA 832

RESULT 12

ABX60639/c

ID ABX60639 standard; cDNA; 2004 BP.

XX AC ABX60639;

XX DT 26-FEB-2003 (first entry)

DE Human musculoskeletal system-associated genomic DNA - SEQ ID 4015.
KW musculoskeletal system; cytostatic; cancer; osteoporosis;
KW gene therapy; vaccine; human; ds.
XX Homo sapiens.
OS XX
PN US2004009488-A1.
XX PD
XX 15-JAN-2004.
XX PF
XX 13-SEP-2002; 2002US-00242515.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0190123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.

```
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764877.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Barash SC;
XX WPI; 2004-090458/09.
XX
XX New nucleic acid molecule, useful for preparing a medicament for
PT preventing, treating or ameliorating a medical condition e.g., cancer of
PT musculoskeletal tissues or osteoporosis.
XX
XX Disclosure; SEQ ID NO 4015; 289pp; English.
XX
XX The invention relates to a novel isolated musculoskeletal system-
CC associated nucleic acid molecule. The nucleic acid of the invention
CC demonstrates cytostatic and osteopathic activities and may be useful for
CC preparing a medicament for preventing, treating or ameliorating a medical
CC condition such as cancer of the musculoskeletal tissues or osteoporosis,
CC possibly via gene therapy or vaccine production. The current sequence is
CC that of the human musculoskeletal system-associated genomic DNA of the
CC invention. The current sequence is not shown within the specification per
CC se but is available on the USPTO web-site
CC http:seqdata.uspto.gov/sequence.html?DocID=20040009488.
XX
XX Sequence 2004 BP; 629 A; 429 C; 353 G; 593 T; 0 U; 0 Other;
SQ
Query Match 77.0%; Score 20.8; DB 12; Length 2004;
Best Local Similarity 91.7%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 GGAGAAATACAAATGGCAAAAGAA 24
Db 855 GGAGAAATACAAAGGACAAAGAA 832
RESULT 14
ADJ31389/c
ID ADJ31389 standard; DNA; 2004 BP.
XX
XX AC ADJ31389;
XX
XX DT 20-MAY-2004 (first entry)
XX
XX DE Human musculoskeletal system-associated genomic DNA - SEQ ID 4016.
XX
XX KW musculoskeletal system; cytostatic; osteopathic; cancer; osteoporosis;
XX gene therapy; vaccine; human; ds.
XX
XX OS Homo sapiens.
XX
XX PN US2004009488-A1.
XX
XX PD 15-JAN-2004.
XX
XX PF 13-SEP-2002; 2002US-00242515.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236389P.
PR 29-SEP-2000; 2000US-0236390P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:51:24 ; Search time 52.4789 Seconds
(without alignments)
841.853 Million cell updates/sec

Title: US-10-039-183A-16
Perfect score: 27
Sequence: 1 ggagaatacaaatggcaagaaag 27

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A-COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B-COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A-COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B-COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS-COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21.2	78.5	601	US-09-949-016-30717	Sequence 30717, A
C 2	21.2	78.5	601	US-09-949-016-30718	Sequence 30718, A
C 3	21.2	78.5	601	US-09-949-016-30719	Sequence 30719, A
C 4	21.2	78.5	601	US-09-949-016-87473	Sequence 87473, A
C 5	21.2	78.5	601	US-09-949-016-87474	Sequence 87474, A
C 6	21.2	78.5	601	US-09-949-016-87475	Sequence 87475, A
C 7	21.2	78.5	107941	US-09-949-016-14206	Sequence 14206, A
C 8	21.2	78.5	109378	US-09-949-016-12391	Sequence 12391, A
C 9	20.8	77.0	96690	US-09-949-016-17103	Sequence 17103, A
C 10	20.6	76.3	496	US-09-270-767-7990	Sequence 7990, Ap
C 11	20.6	76.3	4381	US-09-270-767-23272	Sequence 23272, A
C 12	20.6	76.3	1381	US-09-270-767-3258	Sequence 3258, Ap
C 13	20.6	76.3	1381	US-09-270-767-18540	Sequence 18540, A
C 14	20.4	75.6	511	US-09-270-767-14658	Sequence 14658, A
C 15	20.2	74.8	601	US-09-949-016-76908	Sequence 76908, A
C 16	20.2	74.8	7493	US-08-212-133A-7	Sequence 7, Appl1
C 17	20.2	74.8	7493	US-08-474-503-5	Sequence 5, Appl1
C 18	20.2	74.8	7493	US-08-670-707A-5	Sequence 5, Appl1
C 19	20.2	74.8	7493	US-09-037-601-5	Sequence 5, Appl1
C 20	20.2	74.8	7493	US-09-315-179-5	Sequence 5, Appl1
C 21	20.2	74.8	7493	PCT-US94-13200-5	Sequence 5, Appl1
C 22	20.2	74.8	91665	US-09-949-016-12234	Sequence 12234, A
C 23	20.2	74.8	236341	US-09-949-016-13978	Sequence 13978, A
C 24	19.8	73.3	3008	US-09-503-890-2	Sequence 2, Appl1
C 25	19.8	73.3	19861	US-09-949-016-14498	Sequence 14498, A
C 26	19.6	72.6	601	US-09-949-016-182908	Sequence 182908, A
C 27	19.6	72.6	1140	US-09-134-000C-2622	Sequence 2622, Ap

ALIGNMENTS

RESULT 1

US-09-949-016-30717/c
; Sequence 30717, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30717
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-30717

Query Match 78.5%; Score 21.2; DB 4; Length 601;
Best Local Similarity 88.5%; Pred. No. 42;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGGCAAGAAAA 26
DB 365 GAAGATATACAAATGGCATAGAAAA 340

RESULT 2

US-09-949-016-30718/c
; Sequence 30718, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498

C 28 19.6 72.6 51967 4 US-09-949-016-15982 Sequence 16982, A
C 29 19.6 72.6 101951 4 US-09-949-016-15648 Sequence 15648, A
C 30 19.6 72.6 112112 4 US-09-949-016-15639 Sequence 15639, A
C 31 19.6 72.6 390890 4 US-09-949-016-14720 Sequence 14720, A
C 32 19.2 71.1 219 4 US-09-248-796A-8348 Sequence 8348, Ap
C 33 19.2 71.1 1203 3 US-09-005-298-16 Sequence 16, Appl
C 34 19.2 71.1 1203 3 US-09-005-298-31 Sequence 31, Appl
C 35 19.2 71.1 1203 3 US-08-768-619-16 Sequence 16, Appl
C 36 19.2 71.1 1203 3 US-08-768-619-31 Sequence 31, Appl
C 37 19.2 71.1 1203 5 PCT-US96-09848-16 Sequence 16, Appl
C 38 19.2 71.1 1203 5 PCT-US96-09848-30 Sequence 30, Appl
C 39 19.2 71.1 1306 3 US-09-005-298-14 Sequence 14, Appl
C 40 19.2 71.1 1306 3 US-09-005-298-30 Sequence 30, Appl
C 41 19.2 71.1 1306 3 US-08-768-619-14 Sequence 14, Appl
C 42 19.2 71.1 1306 3 US-08-768-619-30 Sequence 30, Appl
C 43 19.2 71.1 1306 5 PCT-US96-09848-14 Sequence 14, Appl
C 44 19.2 71.1 1306 5 PCT-US96-09848-29 Sequence 29, Appl
C 45 19.2 71.1 2210 1 US-07-710-361-2 Sequence 2, Appl

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30718
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-30718

Query Match      78.5%; Score 21.2; DB 4; Length 601;
Best Local Similarity 88.5%; Pred. No. 42;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGCAAGAAAA 26
Db 225 GAAGATATACAAATGCGCATAGAAAA 200

RESULT 3
US-09-949-016-30719/c
; Sequence 30719, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30719
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-30719

Query Match      78.5%; Score 21.2; DB 4; Length 601;
Best Local Similarity 88.5%; Pred. No. 42;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGCAAGAAAA 26
Db 192 GAAGATATACAAATGCGCATAGAAAA 167

RESULT 4
US-09-949-016-87473/c
; Sequence 87473, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87473
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-87473

Query Match      78.5%; Score 21.2; DB 4; Length 601;
Best Local Similarity 88.5%; Pred. No. 42;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGCAAGAAAA 26
Db 225 GAAGATATACAAATGCGCATAGAAAA 200

RESULT 5
US-09-949-016-87474/c
; Sequence 87474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87474
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-87474

Query Match      78.5%; Score 21.2; DB 4; Length 601;
Best Local Similarity 88.5%; Pred. No. 42;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGCAAGAAAA 26
Db 225 GAAGATATACAAATGCGCATAGAAAA 200

RESULT 6
US-09-949-016-87475/c
; Sequence 87475, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87475
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-87475

Query Match      78.5%; Score 21.2; DB 4; Length 601;
```



```

US-09-949-016-12391

Query Match          78.5%; Score 21.2; DB 4; Length 109378;
Best Local Similarity 88.5%; Pred. No. 76;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GGAGAAATACAAATGCCAAAAGAAAA 26
        ||| ||||| ||||| ||||| |||||
DB      99106 GAAGATATACAAATGGCATGAGAAA 99081

RESULT 9
US-09-949-016-17103
; Sequence 17103, Application US/09949016
; Patent No. 6812339.
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERE
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17103
; LENGTH: 96690
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17103

```

```

Query Match          77.0%; Score 20.8; DB 4; Length 96590;
Best Local Similarity 91.7%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAATACAAATGCGAAAAGAAA 26
    ||||| ||||| ||||| |||||
Db 34238 AGAATTACAGATGCGAAAAGAAA 34261

RESULT 10
US-09-270-767-7990/c
; Sequence 7990, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7990
; LENGTH: 496
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-7990

Query Match          76.3%; Score 20.6; DB 4; Length 496;
Best Local Similarity 85.2%; Pred. No. 69;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGAAAAGAAAAG 27
    ||||| ||||| ||||| |||||
Db 422 GGTGGATACATATGCGAAAAGGAAAG 396

RESULT 11

```

Query Match 76.3%; Score 20.6; DB 4; Length 1381;
Best Local Similarity 85.2%; Pred. No. 78;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 02:04:39 ; Search time 218.282 Seconds
(without alignments)
750.949 Million cell updates/sec

Title: US-10-039-183A-16

Perfect score: 27
Sequence: 1 ggagatacaaatggcaagaagaag 27

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 303525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/FCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/FCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	27	100.0	27	8	US-08-831-310-16
2	27	100.0	27	15	US-10-039-183A-16
3	27	100.0	1448	8	US-08-831-310-3
4	27	100.0	1448	15	US-10-039-183A-3
5	27	100.0	2825	18	US-10-662-126-32
6	22.8	84.4	5914	18	US-10-741-601-5619
7	22.2	82.2	522	18	US-10-425-115-155136
8	21.8	80.7	256190	18	US-10-322-281-320
9	21.4	79.3	82660	13	US-10-087-192-2017
10	21.4	79.3	83493	18	US-10-331-053-61
11	21.2	78.5	378	18	US-10-674-124A-7892
					Sequence 16, Appl
					Sequence 16, Appl
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 32, Appl
					Sequence 5619, Ap
					Sequence 155136,
					Sequence 320, App
					Sequence 2017, Ap
					Sequence 61, Appl
					Sequence 7892, Ap

12	20.8	77.0	699	13	US-10-027-632-17025	Sequence 17025, A
13	20.8	77.0	699	17	US-10-027-632-17025	Sequence 17025, A
c 14	20.8	77.0	2004	9	US-09-764-877-4015	Sequence 4015, Ap
c 15	20.8	77.0	2004	9	US-09-764-877-4015	Sequence 4015, Ap
c 16	20.8	77.0	2004	17	US-10-242-515-4015	Sequence 4015, Ap
c 17	20.8	77.0	2004	17	US-10-242-515-4015	Sequence 4015, Ap
c 18	20.8	77.0	567564	18	US-10-699-156-3	Sequence 3, Appl
c 19	20.6	76.3	668	13	US-10-027-632-121535	Sequence 121535,
c 20	20.6	76.3	668	17	US-10-027-632-121535	Sequence 121535,
c 21	20.6	76.3	3463	13	US-10-027-632-258591	Sequence 258591,
c 22	20.6	76.3	3463	17	US-10-027-632-258591	Sequence 258591,
c 23	20.6	76.3	175590	10	US-09-911-077A-13	Sequence 13, Appl
c 24	20.6	76.3	175590	18	US-10-724-806-13	Sequence 13, Appl
c 25	20.4	75.6	4526	13	US-10-108-605-112	Sequence 112, Appl
c 26	20.4	75.6	1601042	13	US-10-027-632-59064	Sequence 59064, A
c 27	20.4	75.6	1601042	17	US-10-027-632-59064	Sequence 59064, A
c 28	20.2	74.8	635	13	US-10-027-632-221114	Sequence 221114,
c 29	20.2	74.8	635	17	US-10-027-632-221114	Sequence 221114,
c 30	20.2	74.8	2696	18	US-10-437-963-84210	Sequence 84210, A
c 31	20.2	74.8	7493	14	US-10-187-319-5	GENERAL INFORMA
c 32	20.2	74.8	7493	16	US-10-131-510A-5	Sequence 5, Appl
c 33	20.2	74.8	7493	18	US-10-491-464-5	Sequence 5, Appl
c 34	20.2	74.8	7493	19	US-10-938-414-5	GENERAL INFORMA
c 35	20.2	74.8	32183	17	US-10-242-355-1031	Sequence 1031, Ap
c 36	20.2	74.8	45045	19	US-10-741-600-17689	Sequence 17689, A
c 37	19.8	73.3	324	18	US-10-437-963-77574	Sequence 77574, A
c 38	19.8	73.3	2988	18	US-10-437-963-22395	Sequence 22395, A
c 39	19.8	73.3	163701	18	US-10-322-281-439	Sequence 439, App
c 40	19.6	72.6	201	18	US-10-719-993-48777	Sequence 48777, A
c 41	19.6	72.6	201	18	US-10-719-993-48779	Sequence 48779, A
c 42	19.6	72.6	450	17	US-10-424-599-134670	Sequence 134670, A
c 43	19.6	72.6	529	18	US-10-021-323-10810	Sequence 10810, A
c 44	19.6	72.6	712	13	US-10-027-632-17060	Sequence 17060, A
c 45	19.6	72.6	712	17	US-10-027-632-17060	Sequence 17060, A

ALIGNMENTS

RESULT 1
US-08-831-310-16
; Sequence 16, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831.310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-831-310-16

Query Match          100.0%; Score 27; DB 8; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.7; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1 GGAGAAATACAAATGGCAAAAGAAAAG 27
    ||||||||||||||||||||||||||||
Db 1 GGAGAAATACAAATGGCAAAAGAAAAG 27

RESULT 2
US-10-039-183A-16
; Sequence 16, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 27
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-16

Query Match          100.0%; Score 27; DB 15; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.7; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1 GGAGAAATACAAATGGCAAAAGAAAAG 27
    ||||||||||||||||||||||||||||
Db 1 GGAGAAATACAAATGGCAAAAGAAAAG 27

RESULT 3
US-08-831-310-3
; Sequence 3, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,310
```

```
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1448 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 118..1314
; OTHER INFORMATION:
US-08-831-310-3

Query Match          100.0%; Score 27; DB 8; Length 1448;
Best Local Similarity 100.0%; Pred. No. 4.3; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1 GGAGAAATACAAATGGCAAAAGAAAAG 27
    ||||||||||||||||||||||||||||
Db 106 GGAGAAATACAAATGGCAAAAGAAAAG 132

RESULT 4
US-10-039-183A-3
; Sequence 3, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Lissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1448
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (118)...(1314)
US-10-039-183A-3

Query Match          100.0%; Score 27; DB 15; Length 1448;
Best Local Similarity 100.0%; Pred. No. 4.3; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1 GGAGAAATACAAATGGCAAAAGAAAAG 27
    ||||||||||||||||||||||||||||
Db 106 GGAGAAATACAAATGGCAAAAGAAAAG 132

RESULT 5
US-10-662-126-32
; Sequence 32, Application US/10662126
; Publication No. US20050063987A1
```

GENERAL INFORMATION:
APPLICANT: Knapp, Bernhard
APPLICANT: Hundt, Erika
APPLICANT: Schmidt, Karl-Heinz
TITLE OF INVENTION: Proteins, In Particular Membrane Proteins, of
Helicobacter Pylori, Their Preparation and Use
FILE REFERENCE: CHIR-0340
CURRENT APPLICATION NUMBER: US/10/662,126
CURRENT FILING DATE: 2003-09-12
PRIOR APPLICATION NUMBER: 09/230,158
PRIOR FILING DATE: 1999-01-19
PRIOR APPLICATION NUMBER: PCT/IB97/00981
PRIOR FILING DATE: 1997-07-25
NUMBER OF SEQ ID NOS: 52
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 32
LENGTH: 2825
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (891)..(2090)
OTHER INFORMATION: bp protein; 42 kD protein from Helicobacter pylori
US-10-662-126-32

Query Match 100.0%; Score 27; DB 19; Length 2825;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGAAAGAAAAG 27
DB 879 GGAGAAATACAAATGCGAAAGAAAAG 905

RESULT 6
US-10-741-601-5619
Sequence 5619, Application US/10741601
Publication No. US20040166519A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
STENOSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001500
CURRENT APPLICATION NUMBER: US/10/741,601
CURRENT FILING DATE: 2003-12-22
NUMBER OF SEQ ID NOS: 26415
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5619
LENGTH: 59914
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(59914)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5619

Query Match 84.4%; Score 22.8; DB 18; Length 59914;
Best Local Similarity 92.3%; Pred. No. 26+02; 2; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGAAAGAAAAG 26
DB 56532 GGAGAAATAGAAATGCGAAAGAAAAG 56557

RESULT 7
US-10-425-115-155136
Sequence 155136, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 155136
LENGTH: 522
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_73066C.1
US-10-425-115-155136

Query Match 82.2%; Score 22.2; DB 18; Length 522;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGAAAGAAAAG 27
DB 490 GGAGAAATACAAATGCGAAAGAAAAG 516

RESULT 8
US-10-322-281-320/c
Sequence 320, Application US/10322281
Publication No. US20040126762A1
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc S. Malandro
TITLE OF INVENTION: Novel Compositions and Methods in Cancer
FILE REFERENCE: 529452001000
CURRENT APPLICATION NUMBER: US/10/322,281
CURRENT FILING DATE: 2002-12-17
NUMBER OF SEQ ID NOS: 866
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 320
LENGTH: 256190
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(256190)
OTHER INFORMATION: n = A,T,C or G
US-10-322-281-320

Query Match 80.7%; Score 21.8; DB 18; Length 256190;
Best Local Similarity 92.0%; Pred. No. 5.3e+02;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGCGAAAGAAAAG 25
DB 243196 GGAGAAATACAAATGCGAAAGAAAAG 243172

RESULT 9
US-10-087-192-2017
Sequence 2017, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02


```
US-10-027-632-17025
; Sequence 17025, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17025
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-17025

Query Match      77.0%; Score 20.8; DB 17; Length 699;
Best Local Similarity 91.7%; Pred. No. 6.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4  GAAATACAAATGGCAAAAGAAAG 27
      ||||| ||||| ||||| ||||| |||||
Db      36  GAAATAGAAATGGAAAGAAAGAAAG 59

RESULT 14
US-09-764-877-4015/c
; Sequence 4015, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4015
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-4015

Query Match      77.0%; Score 20.8; DB 9; Length 2004;
Best Local Similarity 91.7%; Pred. No. 7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GGAGAAATACAAATGGCAAAAGAA 24
      ||||| ||||| ||||| ||||| |||||
Db      855  GGAGAAATACAAAGGACAAAGAA 832

RESULT 15
US-09-764-877-4016/c
; Sequence 4016, Application US/09764877
; Patent No. US20020147140A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4016
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-4016

Query Match      77.0%; Score 20.8; DB 9; Length 2004;
Best Local Similarity 91.7%; Pred. No. 7e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  GGAGAAATACAAATGGCAAAAGAA 24
      ||||| ||||| ||||| ||||| |||||
Db      855  GGAGAAATACAAAGGACAAAGAA 832

Search completed: April 24, 2005, 05:52:20
Job time : 221.282 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:31:54 ; Search time 1430.24 Seconds
(without alignments)
718.576 Million cell updates/sec

Title: US-10-039-183A-16
Perfect score: 27
Sequence: 1 ggagaaatatacaaatggcaaaagaaaag 27

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.4	86.7	467	AZ734772	AZ734772 RPCI-24-1
C 2	23.4	86.7	567	AQ488891	AQ488891 RPCI-11-2
C 3	22.8	84.4	727	CC727922	CC727922 OGMAU15TV
4	22.8	84.4	780	CC727912	CC727912 OGMAU15TH
5	22.8	84.4	979	CG009054	CG009054 ZUACU15TH
6	22.2	82.2	525	CE522394	CE522394 tigr-gss-
7	22.2	82.2	907	AZ678585	AZ678585 ENTIO09TF
C 8	21.8	80.7	135	AQ488851	AQ488851 RPCI-11-2
9	21.8	80.7	396	AI593036	AI593036 vv04d03.x
C 10	21.8	80.7	590	AQ568710	AQ568710 HS 5319.B
11	21.8	80.7	813	CR319035	CR319035 Medicago
12	21.2	78.5	298	CF309071	CF309071 ABF--03-B
13	21.2	78.5	420	BQ453200	BQ453200 sa097d08.
14	21.2	78.5	442	AZ814634	AZ814634 2M0082021
15	21.2	78.5	470	AI826431	AI826431 wk34b05.x
C 16	21.2	78.5	480	CE307471	CE307471 tigr-gss-
C 17	21.2	78.5	591	CF791611	CF791611 879842 MA
18	21.2	78.5	625	BZ780220	BZ780220 1138b05.b
C 19	21.2	78.5	676	AZ999581	AZ999581 2M0287C01
20	21.2	78.5	795	AZ716242	AZ716242 RPCI-24-1
21	21.2	78.5	808	CG051043	CG051043 PUIEL66TB
22	21.2	78.5	811	BU749906	BU749906 CH34029.F
C 23	21.2	78.5	905	AZ210829	AZ210829 SP.0153.B
C 24	21.2	78.5	938	BU749905	BU749905 CH34029.F

C 25	21.2	78.5	946	7	CF409276	CF409276 CH34059.F
26	21.2	78.5	946	7	CF409277	CF409277 CH34059.F
C 27	21.2	78.5	976	9	CG463316	CG463316 PUFXM63TB
C 28	21.2	78.5	1111	9	AG304869	AG304869 Mus muscu
C 29	21.2	78.5	1201	9	AG305267	AG305267 Mus muscu
30	21.2	78.5	1848	2	BE964937	BE964937 601658711
31	20.8	77.0	394	2	BF372541	BF372541 CM1-PT005
C 32	20.8	77.0	479	5	BY506979	BY506979 BY506979
33	20.8	77.0	509	9	CE647377	CE647377 tigr-gss-
34	20.8	77.0	565	2	BF439598	BF439598 nab65f09.
35	20.8	77.0	591	8	AQ284912	AQ284912 RPCI11-99
C 36	20.8	77.0	596	2	AW768154	AW768154 da70a09.Y
C 37	20.8	77.0	623	8	BZ847869	BZ847869 CH240.232
38	20.8	77.0	667	9	CE661134	CE661134 OR_CBA009
39	20.8	77.0	694	9	CE414844	CE414844 tigr-gss-
40	20.8	77.0	708	5	BE612579	BE612579 BX612579
C 41	20.8	77.0	717	5	BP709346	BP709346 BP709346
42	20.8	77.0	719	4	BM695360	BM695360 UI-B-CQ1-
43	20.8	77.0	719	8	BZ191259	BZ191259 CH230-279
44	20.8	77.0	734	8	AQ384412	AQ384412 RPCI11-13
C 45	20.8	77.0	743	7	CF283199	CF283199 AGENCOURT

ALIGNMENTS

RESULT 1
AZ734772
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AZ734772 467 bp DNA linear GSS 25-JAN-2001
RPCI-24-12004.TJ RPCI-24 Mus musculus genomic clone RPCI-24-12004,
Genomic survey sequence.
AZ734772.1 GI:12500142
GSS.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 467)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levine,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other GSSs: RPCI-24-12004.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
library availability, please contact Pieter de Jong BAC
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 120 row: O column: 4
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1..467
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-12004"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J DNA."

ORIGIN

Query Match 86.7%; Score 23.4; DB 8; Length 467;
Best Local Similarity 96.0%; Pred. No. 4.7e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAATACAAATGGCAAAAGAAA 26

Db 307 GAGAAATACAAATGGCAAAAGAAA 331

RESULT 2

AQ48891/c

LOCUS

DEFINITION RPCI-11-267N1.TV RPCI-11 Homo sapiens genomic clone RPCI-11-267N1, genomic survey sequence.

ACCESSION

VERSION AQ48891

KEYWORDS

SOURCE AQ48891.1 GI:4674765

ORGANISM

Homo sapiens (human)

REFERENCE

AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL

COMMENT Map Building

Other_GSSs: RPCI-11-267N1.TJ

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbs@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: T7

Class: BAC ends.

FEATURES

source

1..567

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="GDB:7602456"

/db_xref="taxon:9606"

/clone="RPCI-11-267N1"

/sex="Male"

/cell_type="Lymphocytes"

/clone_lib="RPCI-11"

/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

ORIGIN

Query Match 86.7%; Score 23.4; DB 8; Length 567;

Best Local Similarity 96.0%; Pred. No. 4.7e+02;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAGAAATACAAATGGCAAAAGAAA 25

Db 470 GAAGAAATACAAATGGCAAAAGAAA 446

RESULT 3

CC727922/c

LOCUS

DEFINITION CC727922 727 bp DNA linear GSS 23-JUN-2003

DEFINITION

OGMAU15TV ZM_0.7_1.5 KB Zea mays genomic clone ZMMBma0331C05, genomic survey sequence.

ACCESSION

VERSION CC727922

KEYWORDS

SOURCE CC727922.1 GI:32146855

ORGANISM

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 727)

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE Consortium for Maize Genomics

JOURNAL Unpublished (2002)

COMMENT Other GSSs: OGMAU15TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: T7

Class: sheared ends.

Location/Qualifiers

1..727

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBma0331C05"

/clone_lib="ZM_0.7_1.5_KB"

/note="Vector: pBCSK; Site_1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match 84.4%; Score 22.8; DB 9; Length 727;

Best Local Similarity 92.3%; Pred. No. 7.6e+02;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAAATACAAATGGCAAAAGAAAAG 27

Db 672 GAGAAATACAAATGGCAATAGAAAAG 647

RESULT 4

CC727912

LOCUS

DEFINITION OGMAU15TH ZM_0.7_1.5 KB Zea mays genomic clone ZMMBma0331C05, genomic survey sequence.

ACCESSION CC727912

VERSION CC727912.1 GI:32146845

KEYWORDS GSS.

SOURCE Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 780)

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE Consortium for Maize Genomics

JOURNAL Unpublished (2002)

COMMENT Other GSSs: OGMAU15TV

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: T7

```

Class: sheared ends.
Location/Qualifiers
1..780
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBM0331C05"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      84.4%; Score 22.8; DB 9; Length 780;
Best Local Similarity 92.3%; Pred. No. 7.6e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAATACAAATGCGCAAAAGAAAG 27
Db 56 GAGAAATAGAAATGCGCAATAGAAAG 81

RESULT 5
CG009054
LOCUS
DEFINITION
ZUACU15TH ZM 3.0 4.0 KB Zea mays genomic clone ZMBP0027D05,
genomic survey sequence.
ACCESSION
CG009054
VERSION
CG009054.1 GI:33881220
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 979)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Reanick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Numborg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: ZUACU15TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..979
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBP0027D05"
/clone_lib="ZM 3.0 4.0 KB"
/notes="Vector: pBCSK; Site 1: HincII; 3-4 kb 'unfiltered'
genomic DNA library"

ORIGIN
Query Match      84.4%; Score 22.8; DB 9; Length 979;
Best Local Similarity 92.3%; Pred. No. 7.6e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAATACAAATGCGCAAAAGAAAG 27
Db 578 GAGAAATAGAAATGCGCAATAGAAAG 603

RESULT 6
CE522394
LOCUS
DEFINITION
tigr-gss-dog-17000365661989 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION
CE522394
VERSION
CE522394.1 GI:36839175
KEYWORDS
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 525)
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
22875432
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: shotgun.
Location/Qualifiers
1..525
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/clone="Site 1: BatXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match      82.2%; Score 22.2; DB 9; Length 525;
Best Local Similarity 88.9%; Pred. No. 1.2e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGAAATACAAATGCGCAAAAGAAAG 27
Db 61 GCAGAAATACAAATGCGCAAAAGAAAG 87

RESULT 7
AZ678585
LOCUS
DEFINITION
ENTIO09TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, genomic survey sequence.
ACCESSION
AZ678585
VERSION
AZ678585.1 GI:11815731
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica
ORGANISM
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 907)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun

```

High quality sequence start: 15
 High quality sequence stop: 853.
 Location/Qualifiers
 1. .907

FEATURES

source
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pBAC3.6; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

ORIGIN

Query Match 82.2%; Score 22.2; DB 8; Length 907;
 Best Local Similarity 88.9%; Pred. No. 1.2e+03;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GGAGAAATACAAATGGCAAAAGAAAG 27
 Db 846 GAAGAAAGACAAATGGCAAAAGAGAG 872

RESULT 8

LOCUS AQ488851/c 135 bp DNA linear GSS 24-APR-1999
 DEFINITION RPCI-11-267L9, TV RPCI-11 Homo sapiens genomic clone RPCI-11-267L9, genomic survey sequence.

ACCESSION AQ488851
 VERSION AQ488851.1 GI:4674725
 KEYWORDS GSS.

SOURCE

ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 135)
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

TITLE

JOURNAL

COMMENT

Unpublished (1997)
 Other GSSs: RPCI-11-267L9.TJ
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
 Seq primer: T7
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1. .135
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="GDB:7602416"
 /db_xref="taxon:9606"

/clone="RPCI-11-267L9"
 /sex="Male"
 /cell_type="Lymphocytes"
 /clone_lib="RPCI-11"
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;
 RPCI11 Human Male BAC Library"

ORIGIN

Query Match 80.7%; Score 21.8; DB 8; Length 135;
 Best Local Similarity 92.0%; Pred. No. 1.7e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGAGAAATACAAATGGCAAAAGAAA 25
 Db 108 GAAGAAAGACAAATGGCAAAAGAAA 84

RESULT 9

LOCUS AI593036 396 bp mRNA linear EST 21-APR-1999
 DEFINITION vv04d03.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone IMAGE:1210661 3', mRNA sequence.

ACCESSION AI593036
 VERSION AI593036
 KEYWORDS EST.

SOURCE

ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 396)
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:647005

This clone was previously sequenced on the 5' end only, this new data is from the 3' end
 Possible reversed clone: polyT not found
 High quality sequence stop: 385.

FEATURES

source

Location/Qualifiers
 1. .396
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NIH Swiss"
 /db_xref="taxon:10090"
 /clone="IMAGE:1210661"
 /sex="pooled"
 /tissue_type="heart"
 /dev_stage="13 day embryos"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratagene mouse heart (#937316)"
 /note="Organ: heart; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 80.7%; Score 21.8; DB 1; Length 396;
 Best Local Similarity 88.5%; Pred. No. 1.7e+03;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 GGAAGATACAAATGCGAAGAAGAAA 26
||| ||||| ||||| ||||| ||||| |||||
Db 330 GGGAATAAAAAATGGCAGAAGAAA 355

RESULT 10
LOCUS HS_5319_B1_A12_T7A_RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=895 Col=23 Row=B, genomic survey sequence.
ACCESSION AQ568710 linear GSS 01-JUN-1999
VERSION HS_5319_B1_A12_T7A_RPCI-11
SOURCE genomic clone Plate=895 Col=23 Row=B, genomic survey sequence.
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 590)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 1049764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pietere@jeng.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 895 row: B column: 23
Seq primer: T7
Class: BAC ends
High quality sequence stop: 590.
FEATURES
source
Location/Qualifiers
1..590
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=895 Col=23 Row=B"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

Query Match 80.7%; Score 21.8; DB 8; Length 590;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAATACAAATGCGAAGAAGAAA 26
||| ||||| ||||| ||||| ||||| |||||
Db 300 GAGAAATACAAATGCGAATTAA 276

RESULT 11
LOCUS CR319035 linear GSS 01-MAR-2004
DEFINITION Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago truncatula, genomic survey sequence.
ACCESSION CR319035
VERSION CR319035.1
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
REFERENCE 1 (bases 1 to 813)
AUTHORS Direct Submission
TITLE Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr
FEATURES
source
Location/Qualifiers
1..813
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="Jemalong A17"
/db_xref="taxon:3880"
/clone_lib="MTE1"
/note="Vector: pIndigoBAC ; Site 1: EcoRI ; Site 2: EcoRI ; Debelle E. and Chalhou B.-Genoscope sequence ID : mcel-44M16RM1"

Query Match 80.7%; Score 21.8; DB 9; Length 813;
Best Local Similarity 92.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AGAATACAAATGCGAAGAAGAAA 27
||||| ||||| ||||| ||||| ||||| |||||
Db 650 AGAATGCAATTCGAAAGAAAAG 674

RESULT 12
LOCUS CF309071 298 bp mRNA EST 15-AUG-2003
DEFINITION ABF--03-B23.g1 ABF3-overexpressing transgenic rice plasmid cDNA library (ABF) Oryza sativa (japonica cultivar-group) cDNA clone ABF--03-B23, mRNA sequence.
ACCESSION CF309071
VERSION CF309071.1 GI:33680832
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 298)
AUTHORS Kim,J.S., Jun,K.M., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C., Song,S.I., Kim,J.K., Kim,Y.-K. and Rahm,B.H.
TITLE Large-scale Sequencing Analysis of Rice ESTs
JOURNAL Unpublished (2003)
CONTACT: Nahm B.H.
COMMENT Genomics and Genetics Institute, Greengene Biotech Inc.; Division of Bioscience and Bioinformatics, Myongji University
Yongin, Kyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnaheggbio.com, bhnaheggbio.ac.kr.
FEATURES
source
Location/Qualifiers
1..298
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="ABF--03-B23"
/tissue_type="leaf"
/dev_stage="14 days after germination"
/lab_host="E.coli DH10B"
/clone_lib="ABF3-overexpressing transgenic rice plasmid"

```



```

/clone="UUCG2M0082O21"
/sex="Male"
/lab host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone lib="Mouse 10kb plasmid UUCG1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```

ORIGIN

```

Query Match      78.5%; Score 21.2; DB 8; Length 442;
Best Local Similarity 88.5%; Pred. NO. 2.7e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 GGAGAAATACAAATGGCAAAAGAAAA 26
Db 88 GGAGAAATACAAATGGCAATTTAAAA 113

```

RESULT 15

```

A1826431
LOCUS
DEFINITION
  wk34b05.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:2417265 3'
  similar to contains element A3R repetitive element ;, mRNA
  sequence.
ACCESSION
  A1826431
VERSION
  A1826431.1 GI:5447102
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 470)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: M. Bento Soares, Ph.D.
  DNA Sequencing by: Greg Lennon, Ph.D.
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Insert Length: 1263 Std Error: 0.00
  Seq primer: -40UP from Gibco
  High quality sequence stop: 417.
  Location/Qualifiers
    1..470
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:2417265"
      /sex="male"
      /tissue_type="normal prostate"

```

FEATURES source

```

/lab host="DH10B"
/clone lib="NCI_CGAP_Pr22"
/notes="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

```

ORIGIN

```

Query Match      78.5%; Score 21.2; DB 1; Length 470;
Best Local Similarity 88.5%; Pred. NO. 2.7e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 1 GGAGAAATACAAATGGCAAAAGAAAA 26

Db 286 GGAGAAATACAAATGGCAAAAGAAAA 311

Search completed: April 24, 2005, 05:28:24

Job time : 1432.24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 17:27:10 ; Search time 700.859 Seconds
(without alignments)
1866.696 Million cell updates/sec

Title: US-10-039-183A-16
Perfect score: 27
Sequence: 1 ggagaaatacaaatggcaaaagaaaag 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ats:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	27	6	BD082354 76 kDa, 3
2	27	100.0	1448	6	BD082347 76 kDa, 3
3	27	100.0	2825	6	BD009865 Proteins,
4	27	100.0	10532	1	AE000626 Helicobac
5	27	100.0	15292	1	AE001541 Helicobac
6	24.4	80.4	113794	2	AC151939 Dasytus n
7	23.4	86.7	163891	9	AC092647 Homo sapi
8	23.4	86.7	206675	10	AC105330 Mus muscu
9	22.8	84.4	118958	2	AC151655 Dasytus n
10	22.8	84.4	118958	2	AC151655 Dasytus n
11	22.8	84.4	138155	2	AC151564 Dasytus n
12	22.8	84.4	142189	2	AC148072 Dasytus n
13	22.8	84.4	148254	2	AC151566 Dasytus n
14	22.8	84.4	164878	2	AC148648 Dasytus n
15	22.8	84.4	173270	9	AC024886 Homo sapi
16	22.8	84.4	328187	2	AC117393 Homo sapi
17	22.4	83.0	115613	2	AC151455 Dasytus n
18	22.4	83.0	239434	2	AC116215 Rattus no
19	22.2	82.2	85615	9	AC134051 Homo sapi

C	20	22.2	82.2	87398	10	AC005855
C	21	22.2	82.2	96288	3	AC005749
C	22	22.2	82.2	105823	2	AC019871 Drosophil
C	23	22.2	82.2	110000	10	AE008684_2
C	24	22.2	82.2	161920	2	AP002502 Homo sapi
C	25	22.2	82.2	167069	2	AC067833 Homo sapi
C	26	22.2	82.2	168922	3	AP002007 Homo sapi
C	27	22.2	82.2	171000	9	AC092244 Drosophil
C	28	22.2	82.2	172859	2	AC010815 Homo sapi
C	29	22.2	82.2	187516	2	AC009648 Homo sapi
C	30	22.2	82.2	337636	3	AE003584 Drosophil
C	31	21.8	80.7	45136	2	AC100349 Mus muscu
C	32	21.8	80.7	92725	9	AC010350 Homo sapi
C	33	21.8	80.7	119736	2	AC108098 Homo sapi
C	34	21.8	80.7	139378	9	HS64K7
C	35	21.8	80.7	144386	2	AC151502 Dasytus n
C	36	21.8	80.7	151250	2	AC151944 Dasytus n
C	37	21.8	80.7	152776	10	AC068064 Mus muscu
C	38	21.8	80.7	160359	9	AC069272 Homo sapi
C	39	21.8	80.7	169663	9	AC146232 Pan trogl
C	40	21.8	80.7	185685	2	AC024283 Homo sapi
C	41	21.8	80.7	192336	9	AL158155 Human DNA
C	42	21.8	80.7	196468	10	AC111065 Mus muscu
C	43	21.8	80.7	200538	10	AC122476 Mus muscu
C	44	21.8	80.7	210300	2	AC137153 Mus muscu
C	45	21.8	80.7	252215	2	AC095632 Rattus no

ALIGNMENTS

RESULT 1
BD082354
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BD082354 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and corresponding polynucleotide molecules.
BD082354.1 GI:22627964
JP 2001523954-A/61.
Mastadenovirus
Viruses; dsDNA viruses, no RNA stage; Adenoviridae.
1 (bases 1 to 27)
Kleanthous, H., Lissolo, L., Tomb, J.F., Miller, C. and Garawi, A.A.
76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and corresponding polynucleotide molecules
Patent: JP 2001523954-A 61 27-NOV-2001;
MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
HUMAN GENOME SCIENCES INC
PN JP 2001523954-A/61
PD 27-NOV-2001
PF 31-MAR-1998 JP 1998541962
PI HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI MILLER,
AMAL AL GARAWI
A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,
A61K39/40,
G01N33/554, G01N33/569
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
1. 27
/organism="Mastadenovirus"
/mol_type="genomic DNA"
/db_xref="taxon:10509"

Query Match 100.0%; Score 27; DB 6; Length 27;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GGAGAAATACAAATGGCAAAAAGAAAG	27
Db	1	GGAGAAATACAAATGGCAAAAAGAAAG	27
RESULT 2			
BD082347			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
Qy	1	GGAGAAATACAAATGGCAAAAAGAAAG	27
Db	106	GGAGAAATACAAATGGCAAAAAGAAAG	132
RESULT 3			
BD009865			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

```

/mol_type="genomic DNA"
/strain="26695"
/db_xref="taxon:85962"
complement(142..519)
/gene="HP1199"
complement(142..519)
/gene="HP1199"
/note="similar to SP:P05392 percent identity: 65.04;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L7/L12 (rpl7/l12)"
/protein_id="AAD08245.1"
/db_xref="GI:2314361"
/translacion="MAISKEVLEYIGLSVLELSELVVMFEKFGVSATPTVAGAA
VNGGAARSEKTEFNVLADSGAEKIKVIVRBITGLGLEAKDATEKTPHVLKEG
VNGEAETIKKLEBVGAKVEVK"
complement(565..1059)
/gene="HP1200"
complement(565..1059)
/gene="HP1200"
/note="similar to SP:P42923 PID:786163 GB:AL009126 percent
identity: 30.39; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L10 (rpl10)"
/protein_id="AAD08246.1"
/db_xref="GI:2314362"
/translacion="WQKHQHQHVKVELVANLKSOFADAKALLICDYKGLSVKLEALR
NKARQGLKVQIKNTLHAHAKETGYGSDLDKETNVLWGGDQIALSKLVDFPQKEH
KDHFLVAGLDFKESVVAHVEAVSKLPSKEELMGMLLSVWTAPARYFTGLDNLURKA
KEN"
complement(1168..1872)
/gene="HP1201"
complement(1168..1872)
/gene="HP1201"
/note="similar to GB:J42023 SP:P44342 PID:1003916
PID:1222450 PID:1204767 percent identity: 51.98;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L1 (rpl1)"
/protein_id="AAD08247.1"
/db_xref="GI:2314363"
/translacion="MAKKVFKRLKLFKIQNDKAYGVEQGVVVKLSAKAFDETVE
VALRIGVDPHADQMVRCAGVLPCHGTGKKVAVPAKDIKODEAKNAGADVVGGDLA
EELKNGRIDFDMVIATPDMMAVGVKGRILPGKGLMPNPKGTGTVMDTAKAVSNKSG
QVNFVDDKGNVHAPIGKASPEEKIKENMLELVKTIINRLKPSAKGKIIRNAALSIT
MSPSVSLDAQBLMDIK"
complement(1917..2342)
/gene="HP1202"
complement(1917..2342)
/gene="HP1202"
/note="similar to SP:P29395 GB:Z11839 PID:48184
GB:AE000512 percent identity: 63.83; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L11 (rpl11)"
/protein_id="AAD08248.1"
/db_xref="GI:2314364"
/translacion="MAKKVGEIKLQIPAGKANPSPVPGPALGORGVNMIFCKAFNE
RTKDMGSPNIPVITVYODKSFPTITKKPPVTDLIKASGVKEGSDNPLKKNIAKLTH
KQVEIAQLKMDLNTSWEAAKIVMSARSMSGEVVD"
complement(2360..2890)
/gene="HP1203"
complement(2360..2890)
/gene="HP1203"
/note="similar to SP:P16921 PID:147801 PID:396321
GB:U00096 PID:1790414 percent identity: 41.04; identified
by sequence similarity; putative"
/codon_start=1

```

```

/transl_table=11
/product="transcription termination factor NusG (nusG)"
/protein_id="AAD08249.1"
/db_xref="GI:2314365"
/translacion="MMDVYAIQTSYSGSQSVKKAJENLANDHNIRDIQIEIIVPTEDI
IEVSKSKTKVTERSLPGVYVFIKVDLDTVLWHKIQSLPRVSRFTGENKKPTPLSEAD
IGHILEKNRNRAAPKPKIPFQEGEVVRVVEGFANFTATVEYDVEHRLKLNVSIFG
RNTPIEILHUSVEKII"
complement(3242..3314)
/gene="trNA-Trp-1"
complement(3242..3314)
/gene="trNA-Trp-1"
/product="trNA-Trp"
/note="codon recognized: UGG"
complement(3354..3512)
/gene="HP1204"
complement(3354..3512)
/gene="HP1204"
/note="similar to SP:P23375 percent identity: 55.10;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L33 (rpl33)"
/protein_id="AAD08255.1"
/db_xref="GI:2314371"
/translacion="MKVKIGLKSCDIEDINYSTTKNAKTNTKLELKKPCPRENKHTL
HKEIKLKS"
complement(3548..4747)
/gene="HP1205"
complement(3548..4747)
/gene="HP1205"
/note="similar to SP:P42482 PID:587590 percent identity:
89.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="translation elongation factor EF-Tu (tufB)"
/protein_id="AAD08250.1"
/db_xref="GI:2314366"
/translacion="MAKEFNRTKPHVNIQTIQVHDKTTLTSAASAVLSKGLAEAM
KLVNIDNAPKEKERTIATSHIETENHYAHVDCPGHADVYKNNITGAOQMDGA
IDVVSADGMPOTREHILLRSQVPHIVVFLANKQMDVDQELLELVEMEVRELLSA
YEPFGDDTPIVAGSALRALEAKAGVNGEKVLKMAEVDAYIPTERDTEKTFLM
PVEDVFSIAGTGTVTGRIGRVGVVGVDEIVIGRPTQKTIVTGVEMFRKELEKGEA
GDVNGVLLRGTKEEVRGVLCPGSIPTKHPEGEIYVLSKEGGHPTPTNRRP
QVYVTTDVTGTSITLPEGVEMVMPQDNVKITVELISPALELGTKFAIREGGRTVGAG
VVSNIIE"
complement(4877..4948)
/gene="trNA-Thr-1"
complement(4877..4948)
/gene="trNA-Thr-1"
/product="trNA-Thr"
/note="codon recognized: ACC"
complement(4982..5055)
/gene="trNA-Gly-2"
complement(4982..5055)
/gene="trNA-Gly-2"
/product="trNA-Gly"
/note="codon recognized: GGA"
complement(5074..5155)
/gene="trNA-Tyr-1"
complement(5074..5155)
/gene="trNA-Tyr-1"
/product="trNA-Tyr"
/note="codon recognized: UAC"
complement(5186..5258)
/gene="trNA-Thr-2"
complement(5186..5258)
/gene="trNA-Thr-2"
/product="trNA-Thr"
/note="codon recognized: ACA"
5448..7184
/gene="HP1206"
5448..7184

```

```
/gene="Hp1206"  
/note="similar to SP:P22638 PID:142020 percent identity:  
26.18; identified by sequence similarity; putative"  
/codon_start=1  
/transl_table=11  
/product="multidrug resistance protein (heta)"  
/protein_id="AAD08251.1"  
/db_xref="GI:2314367"  
/translation="MAKKKHKISILKYFLRSLKQIYMLITFKKRVFLLVMAVFSS
```

```
Query Match 100.0%; Score 27; DB 1; Length 10532;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 CGAGAAATACAAATGCGCAAAAGAAAG 27  
|||||  
Db 4759 CGAGAAATACAAATGCGCAAAAGAAAG 4733
```

```
RESULT 5  
AE001541/c  
LOCUS  
DEFINITION Helicobacter pylori, strain J99 section 102 of 132 of the complete  
genome.  
ACCESSION AE001541 AE001439  
VERSION AE001541.1 GI:4155724  
KEYWORDS  
SOURCE Helicobacter pylori J99  
ORGANISM Helicobacter pylori J99  
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
Helicobacteraceae; Helicobacter.  
1 (bases 1 to 15292)  
Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C.,  
Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G.,  
Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C.,  
Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E.,  
Vovis, G.F. and Trust, T.J.
```

```
TITLE Genomic-sequence comparison of two unrelated isolates of the human  
gastric pathogen Helicobacter pylori  
JOURNAL Nature 397 (6715), 176-180 (1999)  
MEDLINE 99120557  
PUBMED 9923682  
2 (bases 1 to 15292)  
King, B.L., Alm, R.A. and Trust, T.J.  
REFERENCE Direct Submission  
AUTHORS Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney  
JOURNAL Street, Cambridge, MA 02139, USA  
COMMENT Address all correspondence to: hp@arcb.us.astra.com or Richard  
A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,  
MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,  
Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,  
Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics  
Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and  
Diane E. Taylor are with the University of Alberta Department of  
Medical Microbiology and Immunology, Edmonton, Alberta, Canada. T6G  
2H7 and the Canadian Bacterial Diseases Network. All other authors  
are with Astra Research Center Boston, 128 Sidney Street,  
Cambridge, MA, 02139. Putative identifications, sequence  
alignments, and name and sequence search capability are available  
at ARCB's World wide Web site. (URL:  
http://www.astra-boston.com/hpylori).
```

```
FEATURES  
source  
1..15292  
/organism="Helicobacter pylori J99"  
/mol_type="genomic DNA"  
/strain="J99"  
/db_xref="taxon:85963"  
complement(257..415)  
/gene="rpmg"  
/note="synonym: jhpl127"  
complement(257..415)  
/gene="rpmg"  
/note="similar to H. pylori 26695 gene HP1204"
```

```
/codon_start=1  
/transl_table=11  
/product="50S RIBOSOMAL PROTEIN L33"  
/protein_id="AAD06710.1"  
/db_xref="GI:4155730"  
/translation="MKVKGIGKSCDCEDINYSTTNAKTNTKLEKFKCPRENKHTL  
HKEIKLKS"  
complement(462..1661)  
/gene="tufa"  
/note="synonym: jhpl128"  
complement(462..1661)  
/gene="tufa"  
/note="similar to H. pylori 26695 gene HP1205"  
/codon_start=1  
/transl_table=11  
/product="ELONGATION FACTOR TU (EF-TU)"  
/protein_id="AAD06711.1"  
/db_xref="GI:4155731"  
/translation="MAKEKFNRTNPHVNIQGTIGHVYHGKTTLSAAISAVLSKGLAEM  
KDYNDINAPQEKERGITATSHIYEYETETRYAHVDCPGHADVYKMTGAAQMDGA  
ILVSAADGMPQPTREHILLSROGVPHIVFLANKODMVDQELLELMVEMVREILISA  
YEPFGDDTPIVAGSALRALBEAKAGNVGEGEKVKLMAEVDSTIPTPERTEKTFSLM  
PVEDVFSIAGVTGVTGRIERGKVVGVGDEIVIGIRATQKTIVTGVGFRKLEKGEA  
GDNVGLIRGTKEEVERGVLCCKPGSITPHKFEETIYLSKEEGGRHTFFFTNYRP  
QFYVRTTDTVGTGSITLPGVEMVMPGDNVKITVELISPALELGTFAIREGGRTVGAG  
VVSNIIE"  
2365..4101  
/gene="jhpl129"  
2365..4101  
/gene="jhpl129"  
/note="similar to H. pylori 26695 gene HP1206"  
/codon_start=1  
/transl_table=11  
/product="abc transporter, ATP-binding protein"  
/protein_id="AAD06712.1"  
/db_xref="GI:4155732"  
/translation="MAKKKHKIPTLKYLRLSKQIYRLITFKKMIFFLLVMAVFSS  
FVEMSLTLMPEITTLASDPSRALDDKQWVDFEHFSSPVRMLYRPFSCVLGTYLF  
RMFVGVFYLRGRFSNKKAYIQKQLFLQHKNNYLSHLNHLNLSLDLIINNKEGM  
RMSFNALSLTEITVIVFFYDTLITNWKHILVLTFTLLALQIIFIVKKVTVLKKKG  
EMAASKAQTLKVFSPFNKITKLDNHEAHKLFGENSRKADHTBIYSTLQVVP  
RYSIETVGSLLILAVAYILFKYGEARMLPTISYALALYRILPSTVTSYNEIA  
YNQATNVWPKSITVEIDVLPDEFNEKITLQNISPAYKSKHPVLKNFNLTKQKG  
KIALIGHSGGCKSTLADIIMGLTVPKSGEIPIDNTLLTSENRRSRWKKIGYIPONIYL  
FDCTVGNIAFGSAIDKRLIKVCKMAHIYDFLCEHEGLKTQVGEKGLSGGQKQRI  
GIARALNDNPEILVDLDEATSDNETESKIMDEIYQIAKKNLILIVIAHRLSTIERCEV  
IIDSQKHQNLG"  
complement(4110..4778)  
/gene="jhpl130"  
complement(4110..4778)  
/gene="jhpl130"  
/note="similar to H. pylori 26695 gene HP1207"  
/codon_start=1  
/transl_table=11  
/product="putative"  
/protein_id="AAD06713.1"  
/db_xref="GI:4155733"  
/translation="MALEVLWDFDGVIFDSMHLKYEGFKALFQKHGNDSEKGLKQPE  
VYHQSGGIRNEKIQFYNEILKTPAQBEIDALALEFGAILBQKLPDRGHLNSEVM  
AFIDKHQNTYIFHIAAALHSELQVLCFEGITKYPKSVESGSPDKPIIANITOKYA  
YDPSRLMIGDSVNDYESAKANKVAFLGYNSKVLKNLVGQDQYQGYKYLESPKGFDLQW  
FAKE"  
complement(4966..5955)  
/gene="M.Hpy1"  
/note="synonym: jhpl131"  
complement(4966..5955)  
/gene="M.Hpy1"  
/note="similar to H. pylori 26695 gene HP1208"  
/codon_start=1  
/transl_table=11  
/product="TYPE II DNA MODIFICATION ENZYME  
(METHYLTRANSFERASE)"  
/protein_id="AAD06714.1"
```


JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

* 8472 51245: contig of 42774 bp in length
 * 51246 51345: gap of unknown length
 * 51346 113794: contig of 62449 bp in length.

FEATURES

source

Location/Qualifiers
 1. .113794

/organism="Daasybus novemcinctus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9361"
 /clone="VMRC5-357C17"
 /clone_lib="VMRC5"
 /note="BAC resource: <http://bacpac.chori.org/>"
 1. .8371

misc_feature

/note="assembly_fragment
 clone end:T7
 vector_side:left"
 8472..51245

misc_feature

/note="assembly_fragment"

misc_feature

51346..113794
 /note="assembly_fragment
 clone end:SP6
 vector_side:left"

ORIGIN

Query Match 90.4%; Score 24.4; DB 2; Length 113794;

Best Local Similarity 96.2%; Pred. No. 80;

Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GGAGAAATACAAATGGCGAAAGAAAA 26

Db 4224 GAAGAAATACAAATGGCGAAAGAAAA 4199

RESULT 7

AC092647/c

LOCUS

AC092647 Homo sapiens BAC clone RP11-419M24 from 7, complete sequence. PRI 08-OCT-2003

DEFINITION

AC092647 AC027372

ACCESSION

AC092647.2 GI:15638844

VERSION

KEYWORDS

HTG.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 163891)

Sulston,J.E. and Wilson,R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

PUBMED

9847074

2 (bases 1 to 163891)

Abbott,A. and VanBrunt,A.

The sequence of Homo sapiens BAC clone RP11-419M24

Unpublished (2001)

3 (bases 1 to 163891)

Waterston,R.H.

Direct Submission

Submitted (19-JUL-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 163891)

Waterston,R.H.

Direct Submission

Submitted (18-SEP-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 163891)

Waterston,R.

Direct Submission

Submitted (01-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park

```

repeat_region /rpt_family="Alu"
622.679
repeat_region /rpt_family="MaLR"
795.861
repeat_region /rpt_family="L2"
862.1008
repeat_region /rpt_family="CT-rich"
1020.1236
repeat_region /rpt_family="Alu"
1367.1693
repeat_region /rpt_family="Alu"
2068.2090
repeat_region /rpt_family="AT-rich"
2221.2277
repeat_region /rpt_family="(GGGAA)n"
2359.2640
repeat_region /rpt_family="Alu"
2741.2929
repeat_region /rpt_family="MIR"
2963.3280
repeat_region /rpt_family="Alu"
3331.3636
repeat_region /rpt_family="Alu"
3821.3959
repeat_region /rpt_family="MER1_type"
3991.4362
repeat_region /rpt_family="MaLR"
4363.4634
repeat_region /rpt_family="MaLR"
4657.4832
repeat_region /rpt_family="Alu"
4833.5147
repeat_region /rpt_family="MaLR"
5148.5454
repeat_region /rpt_family="Alu"
5455.5723
repeat_region /rpt_family="MaLR"
5735.6088
repeat_region /rpt_family="MaLR"
6299.6593
repeat_region /rpt_family="Alu"
7120.7182
repeat_region /rpt_family="(TAA)n"
7183.7489
repeat_region /rpt_family="Alu"
7490.7624
repeat_region /rpt_family="Alu"
7705.7807
repeat_region /rpt_family="Alu"
7808.7837
repeat_region /rpt_family="(TAA)n"
7838.8007
repeat_region /rpt_family="Alu"
8234.8538
repeat_region /rpt_family="Alu"
8596.8743
repeat_region /rpt_family="Alu"
8961.9263
repeat_region /rpt_family="Alu"
9297.9584
repeat_region /rpt_family="Alu"
9646.9956
repeat_region /rpt_family="Alu"
10023.10319
repeat_region /rpt_family="Alu"
10456.10615
repeat_region /rpt_family="Alu"
10638.10744
snRNA /product="snRNA-U6-related"
10638.10744
repeat_region /rpt_family="U6"
10827.11107
repeat_region /rpt_family="Alu"

```

```

repeat_region 11112..11135
/rpt_family="(TAA)n"
repeat_region 11206..11457
/rpt_family="MER1_type"
repeat_region 11462..11543
/rpt_family="L1"
repeat_region 11544..11825
/rpt_family="Alu"
repeat_region 11826..12192
/rpt_family="L1"
repeat_region 12247..12599
/rpt_family="L1"
repeat_region 12600..12892
/rpt_family="Alu"
repeat_region 12893..12979

Query Match 86.7%; Score 23.4; DB 9; Length 163891;
Best Local Similarity 96.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGAGAAATACAAATGGCAAAAGAAA 25
| | | | | | | | | | | | | | | | | | | | |
Db 31527 GAAGAAATACAAATGGCAAAAGAAA 31503

RESULT 8
AC105330/c 206675 bp DNA linear ROD 26-MAY-2004
LOCUS Mus musculus chromosome 15, clone RP23-281C10, complete sequence.
DEFINITION AC105330
ACCESSION AC105330
VERSION AC105330.8 GI:47679266
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 206675)
Birren,B., Nussbaum,C. and Lander,E.
Mus musculus chromosome 15, clone RP23-281C10
JOURNAL Unpublished
2 (bases 1 to 206675)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vasiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 206675)
Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J.,
Choepe,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D.,

```

Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacClean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrin, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuppback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL Submitted (10-APR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

4 (bases 1 to 206675)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, F., Boguslavsky, L., Bouknight, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeRellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Lui, A., Mabbitt, R., MacClean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrin, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuppback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL Submitted (26-MAY-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On May 26, 2004 this sequence version replaced gi:46358229.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@broad.mit.edu

Project Information

Center project name: L1263

Center Clone name: 281_C10

FEATURES

source

Location/Qualifiers

1..206675

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP23-281C10"

/clone_lib="RPC1-23 Female Mouse BAC"

complement(1..6)

/note="clone boundary"

clone end: SP6

site:EcoRI

16..90

/rpt_family="L1M3"

477..523

/rpt_family=" (TCTA)n"

repeat_region complement(1014..1216)
 rpt_family="Lx2"
 2498..2608
 /rpt_family="ID_B1"
 repeat_region 4347..4379
 /rpt_family=" (T)n"
 repeat_region 4775..4924
 /rpt_family="B4A"
 repeat_region 4979..5015
 /rpt_family=" (CA)n"
 repeat_region 5764..5998
 /rpt_family="RMER19B"
 repeat_region complement(6242..6522)
 /rpt_family="Lx7"
 repeat_region 8595..8760
 /rpt_family="L1"
 repeat_region 8793..8822
 /rpt_family="AT_rich"
 repeat_region 9219..9339
 /rpt_family="Lx"
 repeat_region 9815..9952
 /rpt_family="Lx7"
 repeat_region complement(10006..10391)
 /rpt_family="ORR1B2"
 repeat_region complement(10575..10849)
 /rpt_family="IAPEY_LTR"
 repeat_region complement(10858..11200)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(11273..11796)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(12403..12676)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(13103..13212)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(13218..13833)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(13860..14081)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(14096..14210)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(14362..14523)
 /rpt_family="IAPEY_LTR-int"
 repeat_region 14895..14938
 /rpt_family="AT_rich"
 repeat_region complement(14939..15152)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(15143..15237)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(15238..15391)
 /rpt_family="IAPEY_LTR-int"
 repeat_region complement(15406..15699)
 /rpt_family="IAPEY_LTR"
 repeat_region 15786..15851
 /rpt_family="AT_rich"
 repeat_region 16630..17012
 /rpt_family="WTA"
 repeat_region complement(17388..17607)
 /rpt_family="Lx4"
 repeat_region 17634..18147
 /rpt_family="RMER19B"
 repeat_region 18188..18525
 /rpt_family="RMER19B"
 repeat_region complement(18552..18791)
 /rpt_family="Lx4"
 repeat_region 19093..19157
 /rpt_family=" (TC)n"
 repeat_region 19157..19239
 /rpt_family=" (CA)n"
 repeat_region 19275..19545
 /rpt_family="L1"
 repeat_region 20199..20221
 /rpt_family="AT_rich"
 repeat_region 20865..21092


```

Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoombgri.nih.gov
----- Project Information
Center project name: hth
Center clone name: 394122
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 116414 bases at least Q40
Consensus quality: 116712 bases at least Q30
Consensus quality: 116916 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 118658; sum-of-contigs
Quality coverage: 9.64x in Q20 bases; agarose-fp
Quality coverage: 10.57x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5516: contig of 5516 bp in length
* 5517 5616: gap of unknown length
* 5617 26525: contig of 20909 bp in length
* 26526 26626: gap of unknown length
* 26626 73518: contig of 46893 bp in length
* 73519 73618: gap of unknown length
* 73619 118958: contig of 45340 bp in length.
*
FEATURES
    source
        1. .118958
            /organism="Dasyopus novemcinctus"
            /mol_type="genomic DNA"
            /db_xref="taxon:9361"
            /clone_lib="VMRC5-394122"
            /note="BAC resource: http://bacpac.chori.org/"
        1. .5516
            /note="assembly_fragment"
            clone_end.SP6
            vector_side:right
        5617. .26525
            /note="assembly_fragment"
        26626. .73518
            /note="assembly_fragment"
            clone_end.T7
            vector_side:right
        73619. .118958
            /note="assembly_fragment"

ORIGIN
Query Match      84.4%; Score 22.8; DB 2; Length 118958;
Best Local Similarity 92.3%; Pred. NO. 2.9e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGAGAAATACAAATGGCAAAAGAAAA 26
Db 57426 GAAGAAATACAAATGGCAAAAGAAAA 57401

RESULT 11
AC151564/c
LOCUS          AC151564          138155 bp    DNA     linear   HTG 15-SEP-2004
DEFINITION    Dasyopus novemcinctus clone VMRC5-455010, WORKING DRAFT SEQUENCE, 5
ordered pieces.
ACCESSION    AC151564
VERSION      AC151564.1   GI:52077673
KEYWORDS     HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE       Dasyopus novemcinctus (nine-banded armadillo)

```

```

ORGANISM
Dasyopus novemcinctus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Edentata; Dasypodidae; Dasypus.
1 (bases 1 to 138155)
Antonnellis,A., Ayele,K., Benjamin,B., Blakesley,R.W., Boakye,A.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,H., Daki,N.,
Engle,J., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N.,
Ho,S.-L., Hu,P., Hurl,B., Idol,J.R., Jones,C., Kwong,P., Laric,P.,
Larson,S., Lee-Lin,S.-Q., Legaspi,R., Madden,M., Maduro,Q.L.,
Maduro,V.B., Margulies,E.H., Mastello,C., Maskeri,B., Mcbowell,J.,
Mullikin,J.C., Park,M., Portnoy,M.E., Prasad,A., Puri,O.,
Riddix-Dugue,N., Rosas,B., Schandler,K., Schueler,M.G., Shah,K.,
Sison,C., Stantripong,S., Stephen,E., Thomas,J.W., Thomas,P.J.,
Taipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 138155)
Green,E.D.
Direct Submission
Submitted (15-SEP-2004) NIH Intramural Sequencing Center, 8717
Grovenmont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoombgri.nih.gov
----- Project Information
Center project name: hth
Center clone name: 455010
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 134396 bases at least Q40
Consensus quality: 135460 bases at least Q30
Consensus quality: 135973 bases at least Q20
Insert size: 155000; agarose-fp
Insert size: 137755; sum-of-contigs
Quality coverage: 7.71x in Q20 bases; agarose-fp
Quality coverage: 8.68x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 6946: contig of 6946 bp in length
* 6947 7046: gap of unknown length
* 7047 18919: contig of 11873 bp in length
* 18920 19019: gap of unknown length
* 19020 36242: contig of 17223 bp in length
* 36243 36243: gap of unknown length
* 36243 68490: contig of 32148 bp in length
* 68491 68590: gap of unknown length
* 68591 138155: contig of 69565 bp in length.
*
FEATURES
    source
        1. .138155
            /organism="Dasyopus novemcinctus"
            /mol_type="genomic DNA"
            /db_xref="taxon:9361"
            /clone_lib="VMRC5-455010"
            /note="BAC resource: http://bacpac.chori.org/"
        1. .6946
            /note="assembly_fragment"
        7047. .18919
            /note="assembly_fragment"
        19020. .36242
            /note="assembly_fragment"
        36343. .68490

```

```

/note="assembly_fragment
clone_end:SP6
vector_side:left"
/misc_feature
68591..138155
/note="assembly_fragment
clone_end:T7
vector_side:left"

ORIGIN
Query Match      84.4%; Score 22.8; DB 2; Length 138155;
Best Local Similarity 92.3%; Pred. No. 2.8e+02;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGAGAAATACAATGCAAAAGAAA 26
DB 44904 GAAGAAATACAATGCAAAAGAAA 44879

RESULT 12
AC148072
LOCUS      AC148072      142189 bp      DNA      linear      HTG 20-APR-2004
DEFINITION Dasyopus novemcinctus clone VMRC5-408P13, WORKING DRAFT SEQUENCE, 10
ordered pieces.
ACCESSION   AC148072.2 GI:46409669
VERSION     HTG; HTGS PHASE2; HTGS DRAFT.
KEYWORDS    Dasyopus novemcinctus (nine-banded armadillo)
SOURCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Edentata; Dasyopodidae; Dasyopus.
REFERENCE   1 (bases 1 to 142189)
AUTHORS     Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Daki,N., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B.,
Idol,J.R., Jones,C., Karlins,E., Kim,H., Kwong,P., Lalic,P.,
Larson,S., Lee-Lin,S.-Q., Legaapi,R., Maduro,Q.D., Maduro,V.B.,
Margulies,E.H., Mastello,C., Maskeri,B., McDowell,J.,
Mullikin,J.C., Paguirigan,C., Portnoy,M.E., Prasad,A., Puri,O.,
Reddi-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sison,C.,
Stratipopoulos,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L.,
Wetherby,K.D., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 142189)
Green,E.D.
Direct Submission
Submitted (29-JAN-2004) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 142189)
Green,E.D.
Direct Submission
Submitted (20-APR-2004) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
On Apr 20, 2004 this sequence version replaced gi:41386841.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: fnv
Center clone name: 408P13

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out

gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; version 0.990319
 Assembly program: Phrap; version 0.990319
 Consensus quality: 139461 bases at least Q40
 Consensus quality: 140443 bases at least Q30
 Consensus quality: 140950 bases at least Q20
 Insert size: 155000; agarose-fp
 Insert size: 141289; sum-of-contigs
 Quality coverage: 15.19x in Q20 bases; agarose-fp
 Quality coverage: 16.66x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

1 6130: contig of 6130 bp in length
 6131 6230: gap of unknown length
 6231 16935: contig of 10705 bp in length
 16936 17035: gap of unknown length
 17036 24333: contig of 7298 bp in length
 24334 24433: gap of unknown length
 24434 39533: contig of 15100 bp in length
 39534 39633: gap of unknown length
 39634 49979: contig of 10345 bp in length
 49979 50079: gap of unknown length
 50079 53148: contig of 3070 bp in length
 53149 53248: gap of unknown length
 53249 57094: contig of 3845 bp in length
 57094 57193: gap of unknown length
 57194 61669: contig of 4476 bp in length
 61670 61769: gap of unknown length
 61770 90730: contig of 28961 bp in length
 90731 90830: gap of unknown length
 90831 142189: contig of 51359 bp in length.

FEATURES

	Location/Qualifiers	source
	1..142189	/organism="Dasyopus novemcinctus"
		/mol_type="genomic DNA"
		/db_xref="taxon:9361"
		/clone_lib="VMRC5-408P13"
		/clone_lib="VMRC5"
		/note="BAC resource: http://bacpac.chori.org/"
misc_feature	1..6130	
		/note="assembly_fragment"
		/clone_end:SP6
		/vector_side:left
misc_feature	6231..16935	
		/note="assembly_fragment"
misc_feature	17036..24333	
		/note="assembly_fragment"
misc_feature	24434..39533	
		/note="assembly_fragment"
misc_feature	39634..49978	
		/note="assembly_fragment"
misc_feature	50079..53148	
		/note="assembly_fragment"
misc_feature	53249..57093	
		/note="assembly_fragment"
misc_feature	57194..61669	
		/note="assembly_fragment"
misc_feature	61770..90730	
		/note="assembly_fragment"
misc_feature	90831..142189	
		/note="assembly_fragment"

*		* be preserved.
*	1	7425: contig of 7425 bp in length
*	7426	7525: gap of unknown length

Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Plagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gil, R., Correll, J. H., Guevara, M., Gnaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M.,
Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S.,
Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C.,
Lewis, B., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Louis, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M.,
Mei, G., Metzger, M., Miner, Z., Mitchell, T., Mohabbat, K.,
Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, T., Neal, D.,
Neilson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N.,
Nickerson, E., Nwokoko, S., Ogih, M., Okuwonu, G., Oragunye, N.,
Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L.,
Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M.,
Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S.,
Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E.,
Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A.,
Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C.,
Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L.,
Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S.,
Warren, R., Washington, C., Watlington, S., Williams, G.,
Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y.,
Wu, Y. F., Zhou, J., Zorrilla, S., Naylor, S. L., Weinstein, G. and
Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 173270)
Worley, K.C.

Direct Submission
Submitted (02-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 173270)
Worley, K.C.

Direct Submission
Submitted (01-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 173270)
Worley, K.C.

Direct Submission
Submitted (28-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 173270)
Worley, K.C.

Direct Submission
Submitted (27-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Nov 1, 2001 this sequence version replaced gi:16519437.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST

(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT	
FEATURES	Location/Qualifiers
source	1..173270
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clones="RP11-25K24"
misc_feature	complement(1..2066)
	/note="overlaps bases 1..2066 of clone AC117454"
repeat_region	complement(259..684)
	/rpt_family="MER57A"
repeat_region	1209..1372
	/rpt_family="L2"
repeat_region	complement(2596..2721)
	/rpt_family="AluSq/x"
repeat_region	2722..2897
	/rpt_family="AluSp"
repeat_region	2980..3091
	/rpt_family="MIR"
repeat_region	3781..3979
	/rpt_family="MIR"
repeat_region	complement(4002..4186)
	/rpt_family="MER5A"
repeat_region	4476..4950
	/rpt_family="L1MA9"
repeat_region	7145..7190
	/rpt_family=" (CATATA) n"
repeat_region	8285..8320
	/rpt_family=" (CATA) n"
repeat_region	8380..8400
	/rpt_family="AT-rich"
repeat_region	complement(8489..8520)
	/rpt_family="L2"
repeat_region	complement(8642..8888)
	/rpt_family="MIR"
repeat_region	complement(8972..9105)
	/rpt_family="L2"
repeat_region	10672..11018
	/rpt_family="L1MB7"
repeat_region	11019..11332
	/rpt_family="AluSp"
repeat_region	11333..11418
	/rpt_family="L1MB7"
repeat_region	complement(11419..11707)
	/rpt_family="AluSx"
repeat_region	11708..11761
	/rpt_family="L1MB7"
repeat_region	complement(11847..12113)
	/rpt_family="L1MC5"
repeat_region	12114..12425
	/rpt_family="AluSx"

```

repeat_region 13185..13306
/rpt_family="MERS8B"
repeat_region 13767..13843
/rpt_family=" (CATATA)n"
repeat_region 13844..13985
/rpt_family="tigger3 (Golem)"
repeat_region 13985..14213
/rpt_family="tigger3 (Golem)"
repeat_region complement(14962..15308)
/rpt_family="l2"
repeat_region complement(15513..15816)
/rpt_family="AluSp"
repeat_region 16407..16834
/rpt_family="l2"
repeat_region complement(16868..16995)
/rpt_family="l2"
repeat_region 21831..22127
/rpt_family="AluSq"
repeat_region complement(22172..22439)
/rpt_family="AluJo"
repeat_region 23743..24050
/rpt_family="AluSx"
repeat_region complement(24365..24651)
/rpt_family="AluY"
repeat_region 25278..25416
/rpt_family="MERSA"
repeat_region complement(25417..25595)
/rpt_family="MERSA"
repeat_region 25605..25773
/rpt_family="MERSB"
repeat_region 25838..25906
/rpt_family="MIR"

```

Query Match 84.4%; Score 22.8; DB 9; Length 173270;
 Best Local Similarity 92.3%; Pred. No. 2.7e+02;
 Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 GGAGAAATACAAATGGCAAAAGAAA 26
    |||||
Db 132207 GGAGAAATAGAAATGGCAAAAGAAA 132232

```

Search completed: April 24, 2005, 03:21:54
 Job time : 704.859 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:31:54 ; Search time 1536.18 Seconds
(without alignments)
718.576 Million cell updates/sec

Title: US-10-039-183A-14
Perfect score: 29
Sequence: 1 gtggagacacacataatgaaaaaatatc 29

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_hc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23.2	80.0	529	1 AUI75828	AUI75828
C 2	22.8	78.6	498	6 CB045088	CB045088 NISC GC08
C 3	22.6	77.9	466	6 CB269558	CB269558 1008465 H
C 4	22.6	77.9	678	9 AG124859	AG124859 Pan trogl
C 5	22.6	77.9	840	9 BX165051	BX165051 Danio rer
C 6	22.2	76.6	373	6 CD701672	CD701672 EST18196
C 7	22.2	76.6	534	4 BI493718	BI493718 df104902
C 8	22.2	76.6	582	4 BI493719	BI493719 df104902
C 9	22.2	76.6	816	5 BP507156	BP507156 BP507156
C 10	21.8	75.2	101	8 BZ722055	BZ722055 PUDAP74TD
C 11	21.8	75.2	678	9 CB269558	CB269558 GCDN46TV
C 12	21.8	75.2	795	8 BZ723503	BZ723503 PUDAP03TD
C 13	21.8	75.2	824	9 CG309969	CG309969 ZM9B8034
C 14	21.8	75.2	826	8 BZ959169	BZ959169 PUF1M27TB
C 15	21.8	75.2	831	9 CC697577	CC697577 QJUME74TV
C 16	21.8	75.2	835	8 CC433197	CC433197 FUHUJ32TD
C 17	21.8	75.2	984	8 CC433195	CC433195 FUHUJ32TB
C 18	21.6	74.5	206	9 CE392940	CE392940 tigr-gss-
C 19	21.6	74.5	347	2 BF362498	BF362498 CM0-NN005
C 20	21.6	74.5	434	8 AQ208587	AQ208587 HS 3237 A
C 21	21.6	74.5	509	1 AA439260	AA439260 LD1376-5
C 22	21.6	74.5	561	2 AW943700	AW943700 LD44171.3
C 23	21.6	74.5	581	8 AZ218928	AZ218928 Sheared D
C 24	21.6	74.5	603	9 DR12M10T	AL746669 Danio rer

25	21.6	74.5	614	9 TBR548667	AJ548667 Trypanoso
C 26	21.6	74.5	1107	9 CL032403	CL032403 CH216-34H
C 27	21.4	73.8	660	9 CL194070	CL194070 104_418_1
C 28	21.4	73.8	772	9 CL168109	CL168109 104_365_1
C 29	21.2	73.1	221	2 BB425780	BB425780 BB425780
C 30	21.2	73.1	235	5 BX370495	BX370495 BX370495
C 31	21.2	73.1	343	2 AW897235	AW897235 CM0-NN005
C 32	21.2	73.1	421	5 BY453219	BY453219 BY453219
C 33	21.2	73.1	425	2 AW897234	AW897234 CM0-NN005
C 34	21.2	73.1	452	5 BP021697	BP021697 BP021697
C 35	21.2	73.1	493	9 CL876916	CL876916 abf15c12
C 36	21.2	73.1	691	9 CE281463	CE281463 tigr-gss-
C 37	21.2	73.1	736	9 AG539002	AG539002 Mus muscu
C 38	21.2	73.1	741	5 BU481643	BU481643 603845269
C 39	21.2	73.1	770	9 AG496614	AG496614 Mus muscu
C 40	21.2	73.1	833	5 BU224628	BU224628 603399852
C 41	21	72.4	245	2 BB583000	BB583000 BB583000
C 42	21	72.4	372	7 CO220090	CO220090 WS0108.B2
C 43	21	72.4	378	8 AQ035583	AQ035583 CIT-HSP-2
C 44	21	72.4	419	8 AQ542847	AQ542847 RPCI-11-3
C 45	21	72.4	513	7 CO231736	CO231736 WS0051.B2

ALIGNMENTS

RESULT 1
AUI75828/c
LOCUS AUI75828 529 bp mRNA linear EST 11-JUN-2003
DEFINITION AUI75828 Marsupenaeus japonicus adult Marsupenaeus japonicus CDNA
clone FJAS53 reverse, mRNA sequence.
ACCESSION AUI75828
VERSION AUI75828.1 GI:14622235
KEYWORDS EST.
SOURCE Marsupenaeus japonicus
ORGANISM Marsupenaeus japonicus
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
Penaeidae; Marsupenaeus.
REFERENCE 1 (bases 1 to 529)
Rojtinnakorn, J., Hirono, I., Itami, T., Takahashi, Y. and Aoki, T.
Gene expression in haemocytes of kuruma prawn, *Penaeus japonicus*,
in response to infection with WSSV by EST approach
Fish Shellfish Immunol. 13 (1), 69-83 (2002)
JOURNAL 22189627
MEDLINE
PUBMED 12201653
COMMENT Contact: Takaishi Aoki
Aquatic Biosciences
Tokyo University of Fisheries
Konan 4-5-7, Tokyo, Minato-ku 108-8477, Japan
Tel: 81-3-5463-0689
Fax: 81-3-5463-0690
Email: aoki@tokyo-u-fish.ac.jp.
location/Qualifiers
source
1. 529
/organism="Marsupenaeus japonicus"
/mol_type="mRNA"
/db_xref="taxon:27405"
/clone="PJAS53"
/cell_type="hemocytes"
/dev_stage="adult"
/clone_lib="Marsupenaeus japonicus adult"
/note="Vector: lambda ZAP II"

Query Match 80.0%; Score 23.2; DB 1; Length 529;
Best Local Similarity 89.3%; Pred. No. 7.9e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACATGAAAAAATAT 28

Db 247 GTGTGAACACAAAAATGAAAAAATAT 220

```

RESULT 2
CB045088/c
LOCUS      CB045088      498 bp      mRNA      linear      EST 17-JAN-2003
DEFINITION NISC_gc08d06.x1 NCI_CGAP_Col7 Homo sapiens cDNA clone IMAGE:3217979
3', mRNA sequence.
ACCESSION CB045088
VERSION    CB045088.1 GI:27783375
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE   1 (bases 1 to 498)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
CDNA Library Preparation: The I.M.A.G.E. Consortium/LLNL
cDNA Library Arrayed by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAW7869 row: H column: 12
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers
FEATURES             source
    source
    1..498
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:3217979"
        /tissue_type="juvenile granulosa tumor"
        /lab_host="DH10B"
        /clone_lib="NCI CGAP Col7"
        /note="Organ: Colon; Vector: pCMV-SPORT6; Site 1: SalI;
        Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
        Library constructed by Life Technologies."

ORIGIN
Query Match      78.6%; Score 22.8; DB 6; Length 498;
Best Local Similarity 92.3%; Pred. No. 1.1e+03;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2  TGGAGACACACAAATGAAAAAAATA 27
        ||||| ||||| ||||| ||||| |||||
Db      32 TGGAGACACACAAATGAAAAAAATA 7

RESULT 3
CB269558/c
LOCUS      CB269558      466 bp      mRNA      linear      EST 20-FEB-2003
DEFINITION 1008465 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
CDNA 5', mRNA sequence.
ACCESSION CB269558
VERSION    CB269558.1 GI:28444143
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS    Yang, R.-Z., Shuldiner, A. and Gong, D.-W.
TITLE      EST analysis of human adipose gene expression
JOURNAL    Unpublished (2002)
COMMENT    Contact: Gong Da-Wei
Division of Endocrinology, Diabetes and Nutrition
University of Maryland
660 Redwood St, HH497, Baltimore, MD 21201, USA
Tel: 410 706 1672

```

```

Fax: 410 706 1622
Email: dgong@medicine.umaryland.edu
PCR Primers
FORWARD: CTCGGAGCGCGCATTTGTTGGT
BACKWARD: AATGACTCACTATAGGCGAATTGG
Seq primer: GTTGTACCGGGAATTC.
Location/Qualifiers
FEATURES             source
    source
    1..466
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /sex="Male and Female"
        /tissue_type="Adipose"
        /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
        /note="Vector: lambdaTriplex"

ORIGIN
Query Match      77.9%; Score 22.6; DB 6; Length 466;
Best Local Similarity 86.2%; Pred. No. 1.3e+03;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  GTGAGAACACACAAATGAAAAAATATC 29
        ||||| ||||| ||||| ||||| |||||
Db      98 GAGGAGAACACAAATGAAAAAATATTC 70

RESULT 4
AG124859
LOCUS      AG124859      678 bp      DNA      linear      GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-134P02.F, genomic survey sequence.
ACCESSION AG124859
VERSION    AG124859.1 GI:16654024
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      BAC end sequences of Library PTB
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 678)
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector      : pKS145
R.Site 1    : SacI
R.Site 2    : SacI.
Location/Qualifiers
FEATURES             source
    source
    1..678
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /clone="PTB-134P02.F"
        /sex="male"
        /cell_type="lymphoblast"
        /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match      77.9%; Score 22.6; DB 9; Length 678;
Best Local Similarity 86.2%; Pred. No. 1.3e+03;

```

```

Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACAAATGAAAAAATATC 29
Db 552 GAGGAGAACACACAAATGAAAAAATATTC 580

RESULT 5
BX165051
LOCUS      840 bp      DNA      linear      GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-148G22, genomic survey sequence.
ACCESSION  BX165051
VERSION     BX165051.1  GI:27996584
KEYWORDS   GSS.
SOURCE     Danio rerio (zebrafish)
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 840)
AUTHORS   Humphrey, S.J., Huckle, E. and Durham, J.L.
TITLE     Direct Submission
JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Unpublished
COMMENT   This sequence was generated from the SP6 end of BAC 148G22. 148G22
            is part of the Daniokey BAC Library created by R. Plaaterk and N.V.
            Keygens. Further details:
            http://www.sanger.ac.uk/projects/D_rerio/.

FEATURES             Location/Qualifiers
     source
     1..840
     /organism="Danio rerio"
     /mol_type="genomic DNA"
     /db_xref="taxon:7955"
     /clone="DKEY-148G22"
     /tissue_type="Testis"
     /note="Vector pindigoBAC-536"

ORIGIN
Query Match      77.9%; Score 22.6; DB 9; Length 840;
Best Local Similarity 86.2%; Pred. No. 1.3e+03;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACAAATGAAAAAATATC 29
Db 799 GTGGAGAACACACAAATGAAAAAATATC 827

RESULT 6
CD701672/c
LOCUS      373 bp      mRNA      linear      EST 25-JUN-2003
DEFINITION EST18196 human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION  CD701672
VERSION     CD701672.1  GI:32232302
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 373)
AUTHORS   Liu, X.-Q., Zhou, Y., Zhang, L.-J., Xu, H., Chen, H.-K., Pan, Z.-G. and
            Zeng, Y.-X.
TITLE     Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL   Unpublished (2003)
COMMENT   Contact: Yixin Zeng
            Cancer Center
            Sun Yat-sen University
            651 Dongfeng Road East, Guangzhou 510060, China
            Tel: 86-1380-9770-743
            Fax: 86-20-8775-4506
            Email: yxzeng@gzsums.edu.cn.

FEATURES             Location/Qualifiers
     source
     1..373

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clone_lib="human nasopharynx"
/note="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"

ORIGIN
Query Match      76.6%; Score 22.2; DB 6; Length 373;
Best Local Similarity 88.9%; Pred. No. 1.7e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GGAGAACACACAAATGAAAAAATATC 29
Db 256 GCAGAACACACAAAGGAAAAAATATC 230

RESULT 7
BI493718
LOCUS      534 bp      mRNA      linear      EST 28-AUG-2001
DEFINITION df104902.w1 Morton Fetal Cochlea Homo sapiens cDNA clone
            IMAGE:2541123 3', mRNA sequence.
ACCESSION  BI493718
VERSION     BI493718.1  GI:15333062
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 534)
AUTHORS   Robertson, N.G., Khetarpal, U., Gutierrez-Espeleta, G.A., Bieber, F.R.
            and Morton, C.C.
TITLE     Isolation of novel and known genes from a human fetal cochlear cDNA
            library using subtractive hybridization and differential screening
JOURNAL   Genomics 23, 42-50 (1994)
MEDLINE   95130111
PUBMED    7829101
COMMENT   Contact: Morton, C. C.
            Departments of Pathology and Obstetrics, Gynecology and
            Reproductive Biology
            Brigham and Women's Hospital
            75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
            Tel: 617 732 7980
            Fax: 617 738 6996
            Email: cmorton@rics.bwh.harvard.edu
            DNA sequencing and analyses were performed by National Institutes
            of Health Intramural Sequencing Center (NISC; see
            http://www.nisc.nih.gov).
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Plate: LLAM6329 row: N column: 4
            Seq primer: T7 primer.

FEATURES             Location/Qualifiers
     source
     1..534
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:2541123"
     /tissue_type="cochlea"
     /dev_stage="16-22 week fetus"
     /lab_host="SOLR cells (kanamycin resistant)"
     /clone_lib="Morton Fetal Cochlea"
     /note="Organ: ear; Vector: pBluescript SK-; Site 1: EcoRI;
            Site 2: XhoI; Reference: Genomics 23 42-50 (1994) Cloned.
            37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
            XR Vector. Library constructed by N. Robertson, C. Morton.
            -5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
            sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
Query Match      76.6%; Score 22.2; DB 4; Length 534;

```

Best Local Similarity 88.9%; Pred. No. 1.7e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GGAGAACACACAAATGAAAAAATATC 29
DB 461 GCAGAACACAAAGGAAAAAATATC 487

RESULT 8
BI493719/c
LOCUS
DEFINITION
IMAGE:2541123 5', mRNA sequence.

ACCESSION
BI493719

VERSION
BI493719.1 GI:15333063

KEYWORDS
EST.

ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.

TITLE
Isolation of novel and known genes from a human fetal cochlear cDNA
library using subtractive hybridization and differential screening

JOURNAL
MEDLINE
PUBMED
95310111

COMMENT
Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology

Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980

Fax: 617 738 6996
Email: cmorton@rics.bwh.harvard.edu

DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see

http://www.nisic.nih.gov).

This clone is available royalty-free through LILN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Plate: LLAM6329 row: N column: 4
Seq primer: M13RP1 reverse primer (ABI).

FEATURES
source
Location/Qualifiers

1..582
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2541123"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site 1: EcoRI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGCACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

Query Match 76.6%; Score 22.2; DB 4; Length 582;
Best Local Similarity 88.9%; Pred. No. 1.7e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GGAGAACACACAAATGAAAAAATATC 29

DB 105 GCAGAACACAAAGGAAAAAATATC 79

RESULT 9

BP507156/c

LOCUS
DEFINITION
BP507156 Hydra magnipapillata cDNA library Hydra magnipapillata
cDNA clone hm_03095, mRNA sequence.

ACCESSION
BP507156

VERSION
BP507156.1 GI:34773285

KEYWORDS
EST.

ORGANISM
Hydra magnipapillata

Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
Hydridae; Hydra.

REFERENCE
AUTHORS
Hayakawa,S., Hwang,J.S., Nishimiya-Fujisawa,C., Ogura,A., Ikeo,K.,
Fujisawa,T. and Gojobori,T.

TITLE
Hydra EST project

JOURNAL
Unpublished (2003)

COMMENT
Contact: Jung Shan Hwang
Center for Information Biology and DNA Data Bank of Japan
National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-55-981-6847(ex.6898)

Fax: 81-55-981-6848
Email: jhwang@lab.nig.ac.jp, URL: http://www.cib.nig.ac.jp.

FEATURES
source
Location/Qualifiers

1..816
/organism="Hydra magnipapillata"
/mol_type="mRNA"
/strain="105"
/db_xref="taxon:6085"
/clone="hm_03095"
/tissue_type="whole body"
/dev_stage="adult budding stage"
/clone_lib="Hydra magnipapillata cDNA library"

Query Match 76.6%; Score 22.2; DB 5; Length 816;
Best Local Similarity 88.9%; Pred. No. 1.7e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAGACACACAAATGAAAAAATA 27

DB 627 GTGTAGAACACACAAAGAAAAAAGA 601

RESULT 10
BZ722055/c

LOCUS
DEFINITION
BZ722055 PUBAP74TD ZM_0.6_1.0 KB Zea mays genomic clone ZMMBTa143N04,
genomic survey sequence.

ACCESSION
BZ722055

VERSION
BZ722055.1 GI:28515434

KEYWORDS
GSS.

SOURCE
Zea mays

ORGANISM
Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Renick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.

TITLE
Maize Genomics Consortium
Unpublished (2003)

CONTACT: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843

Fax: 301-838-0208
Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers
1..101
/organism="Zea mays"

```

/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa143N04"
/clone_lib="ZM 0.6 1.0 KB"
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"

ORIGIN
Query Match . 75.2%; Score 21.8; DB 8; Length 101;
Best Local Similarity 92.0%; Pred. No. 2.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 AGACACACAAATGAAAAAATATC 29
Db 98 AGCACACAAATGAAACAAATATC 74

RESULT 11
CC698387/c
LOCUS
DEFINITION OGDN46TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0547H20,
genomic survey sequence.
ACCESSION CC698387
VERSION CC698387.1 GI:32103163
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 678)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
TITLE Query Match
JOURNAL Best Local Similarity
COMMENT Other GSSs: OGDN46TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
1..678
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0547H20"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 75.2%; Score 21.8; DB 9; Length 678;
Best Local Similarity 92.0%; Pred. No. 2.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 AGACACACAAATGAAAAAATATC 29
Db 259 AGCACACAAATGAAACAAATATC 235

RESULT 12
BZ723503/c
LOCUS
DEFINITION PUDAP03TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTa143B06,
genomic survey sequence.
ACCESSION BZ723503

```

```

VERSION BZ723503.1 GI:28518673
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 795)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
TITLE Query Match
JOURNAL Best Local Similarity
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
1..795
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa143B06"
/clone_lib="ZM 0.6 1.0 KB"
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"

ORIGIN
Query Match 75.2%; Score 21.8; DB 8; Length 795;
Best Local Similarity 92.0%; Pred. No. 2.4e+03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 AGAACACAAATGAAAAAATATC 29
Db 194 AGCACACAAATGAAACAAATATC 170

RESULT 13
CG909969
LOCUS
DEFINITION ZMMBBb0345N12.r ZMMBBb Zea mays genomic clone ZMMBBb0345N12 3',
genomic survey sequence.
ACCESSION CG909969
VERSION CG909969.1 GI:39609239
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 824)
Yu,Y., Kim,H.R., Hatfield,J., Soderlund,C., Bharti,A.K., Messing,J.
and Wing,R.
Sequencing of the maize genome
Unpublished (2003)
TITLE Query Match
JOURNAL Best Local Similarity
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 480A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: T7
BACKWARD: M13r
Plate: 0345 row: N column: 12
Seq primer: M13r

```

```

FEATURES
  source
    Class: BAC ends.
    Location/Qualifiers
      1..824
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /cultivar="B73"
        /db_xref="taxon:4577"
        /clone="ZMMBB0345N12"
        /lab_host="DH10B"
        /clone_lib="ZMMBB"
        /note="Vector: pBeloBAC11; Site 1: HindIII; Site 2: HindIII; Zea mays L. esp. mays"

ORIGIN
  Query Match      75.2%; Score 21.8; DB 9; Length 824;
  Best Local Similarity 92.0%; Pred. No. 2.4e+03;
  Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  5 AGAACACACAATGAAAAAATATC 29
    |||||||||||||||||||
DB  178 AGCACACACATGAAACAAATATC 202

RESULT 14
BZ959169/c
LOCUS
DEFINITION
  BZ959169 826 bp DNA linear GSS 25-MAR-2003
  PUFIM27TB_ZM_0.6_1.0_KB Zea mays genomic clone ZMMBT351E05,
  genomic survey sequence.
ACCESSION
  BZ959169
VERSION
  BZ959169.1 GI:29174261
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 826)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
  Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other GSSs: OGUME74TH
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TP
  Class: sheared ends.
  Location/Qualifiers
    1..826
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBT351E05"
      /clone_lib="ZM 0.6 1.0 KB"
      /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
      Cot selected genomic DNA library"

ORIGIN
  Query Match      75.2%; Score 21.8; DB 8; Length 826;
  Best Local Similarity 92.0%; Pred. No. 2.4e+03;
  Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  5 AGAACACACAATGAAAAAATATC 29
    |||||||||||||||||||
DB  132 AGCACACACAATGAAACAAATATC 108

RESULT 15

```

```

CC697577
LOCUS
DEFINITION
  CC697577 831 bp DNA linear GSS 19-JUN-2003
  OGUME74TV_ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMA0473M03,
  genomic survey sequence.
ACCESSION
  CC697577
VERSION
  CC697577.1 GI:32102353
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
  1 (bases 1 to 831)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
  Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Other GSSs: OGUME74TH
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TP
  Class: sheared ends.
  Location/Qualifiers
    1..831
      /organism="Zea mays"
      /mol_type="genomic DNA"
      /strain="B73"
      /db_xref="taxon:4577"
      /clone="ZMMBMA0473M03"
      /clone_lib="ZM 0.7 1.5 KB"
      /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
      methylation filtered genomic DNA library"

ORIGIN
  Query Match      75.2%; Score 21.8; DB 9; Length 831;
  Best Local Similarity 92.0%; Pred. No. 2.4e+03;
  Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  5 AGAACACACAATGAAAAAATATC 29
    |||||||||||||||||||
DB  539 AGCACACACAATGAAACAAATATC 563

Search completed: April 24, 2005, 05:28:18
Job time : 1549.18 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 02:04:39 ; Search time 234.451 Seconds
(without alignments)
750.949 Million cell updates/sec

Title: US-10-039-183A-14

Perfect score: 29
Sequence: 1 gtagagacacacaaatgaaaaaaatc 29

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 303525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	29	8	US-08-831-310-14
2	29	100.0	29	15	US-10-039-183A-14
3	29	100.0	1000	9	US-09-881-752A-243
4	29	100.0	1149	8	US-08-831-310-1
5	29	100.0	1149	15	US-10-039-183A-1
6	27.4	94.5	456	17	US-10-335-977-481
7	23.2	80.0	4134	9	US-09-070-927A-541
8	22.4	77.2	421609	18	US-10-367-094-122
9	22.2	76.6	4746	18	US-10-489-242-7
10	22.2	76.6	96589	17	US-10-052-482-214
11	21.6	74.5	1503841	9	US-09-795-668-1

12	21.6	74.5	1503841	9	US-09-795-668-1	Sequence 1, Appli
13	21.6	74.5	1503841	9	US-09-795-668-1	Sequence 1, Appli
c 14	21.2	73.1	523	17	US-10-424-599-129489	Sequence 129489, A
15	21	72.4	1385	17	US-10-425-114-20083	Sequence 20083, A
16	21	72.4	1592	17	US-10-424-599-60506	Sequence 60506, A
17	21	72.4	1766	17	US-10-425-114-8625	Sequence 8625, AP
18	21	72.4	1926	17	US-10-424-599-60505	Sequence 60505, A
c 19	20.6	71.0	3902	13	US-10-108-605-88	Sequence 88, Appli
20	20.6	71.0	4278	17	US-10-282-122A-21174	Sequence 21174, A
21	20.6	71.0	11739	9	US-09-070-927A-150	Sequence 150, App
c 22	20.4	70.3	300000	15	US-10-774-355A-1204	Sequence 1204, Ap
23	20.4	70.3	300000	15	US-10-774-355A-1204	Sequence 1204, Ap
24	20.4	70.3	300000	18	US-10-703-210-33	Sequence 33, Appl
c 25	20.2	69.7	1099	17	US-10-425-114-9557	Sequence 9557, Ap
c 26	20.2	69.7	1101	17	US-10-424-599-133941	Sequence 133941, A
27	20.2	69.7	299598	18	US-10-322-696-16	Sequence 16, Appl
28	20.2	69.7	326014	9	US-09-731-231A-3	Sequence 3, Appli
29	20.2	69.7	326014	18	US-10-751-985-3	Sequence 3, Appli
c 30	20.2	69.7	3673778	16	US-10-312-841-2	Sequence 2, Appli
31	20	69.0	686	13	US-10-027-632-8809	Sequence 8809, Ap
32	20	69.0	686	17	US-10-027-632-8809	Sequence 8809, Ap
c 33	20	69.0	750	17	US-10-424-599-120464	Sequence 120464, A
34	20	69.0	1103	17	US-10-425-114-8260	Sequence 8260, Ap
35	20	69.0	1192	9	US-09-443-704-27	Sequence 27, Appl
36	20	69.0	1192	13	US-10-008-118A-27	Sequence 27, Appl
37	20	69.0	1635	17	US-10-320-797-2242	Sequence 2242, Ap
c 38	20	69.0	1780	14	US-10-198-846-10696	Sequence 10696, A
39	20	69.0	1992	17	US-10-320-797-1242	Sequence 1242, Ap
c 40	20	69.0	2651	17	US-10-104-047-984	Sequence 984, App
c 41	20	69.0	3202	15	US-10-106-698-387	Sequence 387, App
c 42	20	69.0	3346	18	US-10-425-115-175386	Sequence 175386, A
c 43	20	69.0	3401	18	US-10-437-963-16892	Sequence 16892, A
c 44	20	69.0	3992	17	US-10-320-797-242	Sequence 242, App
c 45	20	69.0	4500	17	US-10-221-596B-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-831-310-14
; Sequence 14, Application US/08831310
; Publication No. US2002026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831.310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-831-310-14
Query Match 100.0%; Score 29; DB 8; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
    ||||||||||||||||||||||||||||
Db 1 GTGGAGAACACACAAATGAAAAAATATC 29

RESULT 2
US-10-039-183A-14
; Sequence 14, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Liessolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Helicobacter GHPO 1360 and GHPO 750
; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 29
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-10-039-183A-14

Query Match 100.0%; Score 29; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
    ||||||||||||||||||||||||||||
Db 1 GTGGAGAACACACAAATGAAAAAATATC 29

RESULT 3
US-09-881-752A-243
; Sequence 243, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 243
; LENGTH: 1000
; TYPE: DNA

; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (51)...(947)
US-09-881-752A-243

Query Match 100.0%; Score 29; DB 9; Length 1000;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
    ||||||||||||||||||||||||||||
Db 37 GTGGAGAACACACAAATGAAAAAATATC 65

RESULT 4
US-08-831-310-1
; Sequence 1, Application US/08831310
; Publication No. US20020026035A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold et al.
; TITLE OF INVENTION: Helicobacter GHPO 1360 and
; TITLE OF INVENTION: GHPO 750 Polypeptides and Corresponding Polynucleotides
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,310
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/037001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 106...1002
; OTHER INFORMATION:
; NAME/KEY: Signal Sequence
; LOCATION: 106...166
; OTHER INFORMATION:
US-08-831-310-1

Query Match 100.0%; Score 29; DB 8; Length 1149;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
    ||||||||||||||||||||||||||||
Db 92 GTGGAGAACACACAAATGAAAAAATATC 120
```



```
RESULT 5
US-10-039-183A-1
; Sequence 1, Application US/10039183A
; Publication No. US20030143242A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Liesolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: Polypeptides and Corresponding Polynucleotide Molecules
; FILE REFERENCE: 06132/037002
; CURRENT APPLICATION NUMBER: US/10/039,183A
; PRIORITY FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: US 08/831,310
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (106)...(1002)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (106)...(166)
US-10-039-183A-1

Query Match      100.0%; Score 29; DB 15; Length 1149;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GTGGAGAACACACAAATGAAAAAATATC 29
Db      92 GTGGAGAACACACAAATGAAAAAATATC 120

RESULT 6
US-10-335-977-481
; Sequence 481, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:

US-10-039-183A-14.rnpb
```

```
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 481:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...456
; SEQUENCE DESCRIPTION: SEQ ID NO: 481:
US-10-335-977-481

Query Match      94.5%; Score 27.4; DB 17; Length 456;
Best Local Similarity 96.6%; Pred. No. 3.3;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GTGGAGAACACACAAATGAAAAAATATC 29
Db      2 GTGGAGAACACACAAATGAAAAAATATC 30

RESULT 7
US-09-070-927A-541/c
; Sequence 541, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 541:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
;
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 541:
US-09-070-927A-541

Query Match      80.0%; Score 23.2; DB 9; Length 4134;
Best Local Similarity 89.3%; Pred. No. 1.3e+02;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 TGGAGACACACAATGAAAAAATATC 29
      ||||| ||||| ||||| ||||| |||||
DB      1851 TGGAGAGCAACAATGAAAAAATATC 1824

RESULT 8
US-10-367-094-122
; Sequence 122, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367,094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 122
; LENGTH: 421609
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(421609)
; OTHER INFORMATION: n = A,T,C or G
US-10-367-094-122

Query Match      77.2%; Score 22.4; DB 18; Length 421609;
Best Local Similarity 95.8%; Pred. No. 4e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      6 GAACACACAATGAAAAAATATC 29
      ||||| ||||| ||||| ||||| |||||
DB      319593 GAACACACACTGAAAAAATATC 319616

RESULT 9
US-10-489-242-7
; Sequence 7, Application US/10489242
; Publication No. US20040253710A1
; GENERAL INFORMATION:
; APPLICANT: ACAMBI'S RESEARCH LIMITED
; TITLE OF INVENTION: BACTERIAL VACCINE
; FILE REFERENCE: N.84948A
; CURRENT APPLICATION NUMBER: US/10/489,242
; CURRENT FILING DATE: 2004-03-11
; PRIOR APPLICATION NUMBER: GB 0121998.9
; PRIOR FILING DATE: 2001-09-11
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 4746
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Gene
; LOCATION: (378)..(1103)
; OTHER INFORMATION: csta
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1858)..(3579)
; OTHER INFORMATION: cstaB
; FEATURE:
; NAME/KEY: gene
```

```
; LOCATION: (2266)..(3579)
; OTHER INFORMATION: cstaC
; FEATURE:
; NAME/KEY: gene
; LOCATION: (2668)..(3579)
; OTHER INFORMATION: cstaD
; FEATURE:
; NAME/KEY: gene
; LOCATION: (3031)..(3579)
; OTHER INFORMATION: cstaE
; FEATURE:
; NAME/KEY: gene
; LOCATION: (3577)..(3579)
; OTHER INFORMATION: cstaF
; FEATURE:
; NAME/KEY: gene
; LOCATION: (4153)..(4659)
; OTHER INFORMATION: cstaH
; OTHER INFORMATION: cstaH
US-10-489-242-7

Query Match      76.6%; Score 22.2; DB 18; Length 4746;
Best Local Similarity 88.9%; Pred. No. 2.9e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 TGGAGACACACAATGAAAAAATAT 28
      ||||| ||||| ||||| ||||| |||||
DB      3631 TGGAGACACACACTGAAAAAATGAT 3657

RESULT 10
US-10-052-482-214/c
; Sequence 214, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 214
; LENGTH: 96589
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9857)..(9876)
; OTHER INFORMATION: "T" at positions 9857 to 9876 can be any base
US-10-052-482-214

Query Match      76.6%; Score 22.2; DB 17; Length 96589;
Best Local Similarity 88.9%; Pred. No. 4e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 GGAGAACACACAATGAAAAAATATC 29
      ||||| ||||| ||||| ||||| |||||
DB      40453 GCAGAACACAAAAGGAAAAAATATC 40427

RESULT 11
US-09-795-668-1
; Sequence 1, Application US/09795668
; Patent No. US20020045577A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: r=g or a
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: y=t/u or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k=g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s=g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w=a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b=g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d=a or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h=a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: n=a or g or c or t/u
; US-09-795-668-1

Query Match 74.5%; Score 21.6; DB 9; Length 1503841;
Best Local Similarity 85.7%; Pred. No. 7.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TGGAGACACACATGAAAAAATATC 29
|||||
Db 167284 TGGAGACACACATGAAAAAAGAAAC 167311

RESULT 12
US-09-795-686-1
; Sequence 1, Application US/09/795666
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
```

```
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: r=g or a
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: y=t/u or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k=g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s=g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w=a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b=g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d=a or g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h=a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: n=a or g or c or t/u
; US-09-795-686-1

Query Match 74.5%; Score 21.6; DB 9; Length 1503841;
Best Local Similarity 85.7%; Pred. No. 7.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TGGAGACACACATGAAAAAATATC 29
|||||
Db 167284 TGGAGACACACATGAAAAAAGAAAC 167311

RESULT 13
US-09-946-807-1
; Sequence 1, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: r=g or a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: y=/u or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k=g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w=a or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b=g or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d=a or g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h=a or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v=a or g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: n=a or g or c or t/u
US-09-946-807-1
```

```
Query Match 74.5%; Score 21.6; DB 9; Length 1503841;
Best Local Similarity 85.7%; Pred. No. 7.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 TGGAGACACACAATGAAAAAAATATC 29
||||| ||||| ||||| ||||| |||||
DB 167284 TGGAGACACACAATGAAAAAGAAAC 167311
```

```
RESULT 14
US-10-424-599-129489/c
; Sequence 129489, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223) B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 129489
```

```
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87934C.1
US-10-424-599-129489
```

```
Query Match 73.1%; Score 21.2; DB 17; Length 523;
Best Local Similarity 88.5%; Pred. No. 5.1e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 TGGAGACACACAATGAAAAAAATA 27
||||| ||||| ||||| ||||| |||||
DB 422 TGGAGACAGAAATGAAAAAAATTA 397
```

RESULT 15

```
US-10-425-114-20083
; Sequence 20083, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313) B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 20083
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3170-088-C1_FLI
US-10-425-114-20083
```

```
Query Match 72.4%; Score 21; DB 17; Length 1385;
Best Local Similarity 82.8%; Pred. No. 6.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 GTGGAGACACACAATGAAAAAAATATC 29
||||| ||||| ||||| ||||| |||||
DB 260 GGGGAGACACACAATGAAATAAAGAATC 288
```

```
Search completed: April 24, 2005, 05:52:14
Job time : 242.451 secs
```

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2005, 00:51:24 ; Search time 56.3662 Seconds
(without alignments)
841.853 Million cell updates/sec

Title: US-10-039-183A-14

Perfect score: 29

Sequence: 1 GTGAGAACACACAAATGAAAAAATATC 29

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	1149	US-09-336-115C-23	Sequence 23, Appl
C 2	22.2	76.6	136264	US-09-949-016-12756	Sequence 12756, A
C 3	22.2	76.6	136264	US-09-949-016-13001	Sequence 13001, A
4	21.2	73.1	186734	US-09-949-016-14870	Sequence 14870, A
5	21.2	73.1	193689	US-09-949-016-12350	Sequence 12350, A
6	21.2	73.1	193689	US-09-949-016-13088	Sequence 13088, A
C 7	20.6	71.0	3900	US-08-123-343A-5	Sequence 6, Appl
8	20.6	71.0	4320	US-09-134-000C-2728	Sequence 2728, Ap
C 9	20.2	69.7	69642	US-09-949-016-15339	Sequence 15339, A
10	20	69.0	876	US-09-107-532A-3580	Sequence 3580, Ap
11	20	69.0	236474	US-09-949-016-13418	Sequence 13418, A
12	19.8	68.3	267	US-09-248-796A-10040	Sequence 10040, A
13	19.6	67.6	501	US-09-270-767-29035	Sequence 29035, A
C 14	19.6	67.6	601	US-09-949-016-93939	Sequence 93939, A
15	19.6	67.6	882	US-09-270-767-13134	Sequence 13134, A
C 16	19.6	67.6	1209	US-09-533-029-85	Sequence 85, Appl
17	19.6	67.6	676	US-09-949-016-14855	Sequence 14855, A
C 18	19.6	67.6	200663	US-09-949-016-12569	Sequence 12569, A
C 19	19.6	67.6	203093	US-09-949-016-14445	Sequence 14445, A
20	19.6	67.6	331814	US-09-949-016-12008	Sequence 12008, A
21	19.6	67.6	331814	US-09-949-016-17056	Sequence 17056, A
22	19.4	66.9	601	US-09-949-016-151739	Sequence 151739, A
C 23	19.4	66.9	1492	US-09-774-528-297	Sequence 297, App
C 24	19.4	66.9	317366	US-09-949-016-16001	Sequence 16001, A
C 25	19.4	66.9	421491	US-09-949-016-12805	Sequence 12805, A
C 26	19.4	66.9	421494	US-09-949-016-14060	Sequence 14060, A
27	19.2	66.2	237	US-09-489-039A-378	Sequence 378, App

C 28 19.2 66.2 2824 4 US-09-662-831-4 Sequence 4, Appl
29 19.2 66.2 7186 4 US-09-978-594-1017 Sequence 1017, Ap
30 19.2 66.2 37475 4 US-09-949-016-12299 Sequence 12299, A
31 19.2 66.2 37492 4 US-09-949-016-13198 Sequence 13198, A
32 19.2 66.2 421491 4 US-09-949-016-12805 Sequence 12805, A
33 19.2 66.2 421494 4 US-09-949-016-14060 Sequence 14060, A
34 19 65.5 227 4 US-09-023-655-394 Sequence 394, App
35 19 65.5 601 4 US-09-949-016-161680 Sequence 161680, App
C 36 19 65.5 601 4 US-09-949-016-161681 Sequence 161681, App
37 19 65.5 624 4 US-09-328-352-1345 Sequence 1345, Ap
38 19 65.5 1195 4 US-09-270-767-11635 Sequence 11635, A
39 19 65.5 4982 4 US-09-949-016-14874 Sequence 14874, A
40 19 65.5 14636 3 US-09-173-914-6 Sequence 6, Appl
41 19 65.5 40465 4 US-09-949-016-12561 Sequence 12561, A
42 19 65.5 41125 4 US-09-949-016-12413 Sequence 12413, A
43 19 65.5 41126 4 US-09-949-016-17273 Sequence 17273, A
C 44 19 65.5 54531 4 US-09-949-016-16267 Sequence 16267, A
C 45 19 65.5 161900 4 US-09-949-016-12685 Sequence 12685, A

ALIGNMENTS

RESULT 1

US-09-336-115C-23
; Sequence 23, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (106)...(1002)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (106)...(166)
US-09-336-115C-23

Query Match 100.0%; Score 29; DB 4; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0.2; Indels 0; Gaps 0;
Matches 29; Conservative 0; Mismatches 0;

QY 1 GTGAGAACACACAAATGAAAAAATATC 29
Db 92 GTGAGAACACACAAATGAAAAAATATC 120

RESULT 2

US-09-949-016-12756/c
; Sequence 12756, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12756
; LENGTH: 136264
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(136264)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12756
```

```
Query Match 76.6%; Score 22.2; DB 4; Length 136264;
Best Local Similarity 88.9%; Pred. No. 71;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 GGAGAACACACAAATGAAAAAATATC 29
Db 32453 GCAGAACACAAAGGAAAAAATATC 32427
```

```
RESULT 3
US-09-949-016-13001/C
; Sequence 13001, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13001
; LENGTH: 136265
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(136265)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13001
```

```
Query Match 76.6%; Score 22.2; DB 4; Length 136265;
Best Local Similarity 88.9%; Pred. No. 71;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 3 GGAGAACACACAAATGAAAAAATATC 29
Db 32453 GCAGAACACAAAGGAAAAAATATC 32427
```

```
RESULT 4
US-09-949-016-14870
; Sequence 14870, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14870
; LENGTH: 186734
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14870
```

```
Query Match 73.1%; Score 21.2; DB 4; Length 186734;
Best Local Similarity 88.5%; Pred. No. 1.6e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 4 GAGAACACACAAATGAAAAAATATC 29
Db 104097 GAGAAACACAAATGAAATATGTC 104122
```

```
RESULT 5
US-09-949-016-12350
; Sequence 12350, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12350
; LENGTH: 193689
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12350
```

```
Query Match 73.1%; Score 21.2; DB 4; Length 193689;
Best Local Similarity 88.5%; Pred. No. 1.6e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 4 GAGAACACACAAATGAAAAAATATC 29
Db 104097 GAGAAACACAAATGAAATATGTC 104122
```

```
RESULT 6
US-09-949-016-13088
; Sequence 13088, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

RESULT 10
US-09-107-532A-3580
; Sequence 3580, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm
; TITLE OF INVENTION: NUCLEIC ACID

QY 4 GAGAACACACATGAAAAAATATC 29
 |||||
 DB 171 GACAACATAGATGAATAAATATC 196

RESULT 14
 US-09-949-016-93939/C
 ; Sequence 93939, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 93939
 ; LENGTH: 601
 ; TYPE: DNA
 ; ORGANISM: Human
 US-09-949-016-93939

Query Match 67.6%; Score 19.6; DB 4; Length 601;
 Best Local Similarity 84.6%; Pred. No. 3.7e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 TGGAGAACACACATGAAAAAATA 27
 |||||
 DB 344 TGGGACACACAAATGAATGAATA 319

RESULT 15
 US-09-270-767-13134
 ; Sequence 13134, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 13134
 ; LENGTH: 882
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-13134

Query Match 67.6%; Score 19.6; DB 4; Length 882;
 Best Local Similarity 84.6%; Pred. No. 3.9e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 4 GAGAACACACATGAAAAAATATC 29
 |||||
 DB 552 GACAACATAGATGAATAAATATC 577

Search completed: April 24, 2005, 05:32:58
 Job time : 63.5662 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 10:03:10 ; Search time 190.542 Seconds
(without alignments)
900.968 Million cell updates/sec

Title: US-10-039-183A-14

Perfect score: 29
Sequence: 1 gtgagacacacacataaataatc 29

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	29	AAV07969	Helicobac
2	29	100.0	1000	AAx14041	H. pylori
3	29	100.0	1082	AAV90844	Nucleotid
4	29	100.0	1082	AAV90591	PA
5	29	100.0	1117	AAV90580	Nucleotid
6	29	100.0	1149	AAV07963	Helicobac
7	29	100.0	1149	AAV07963	Helicobac
8	27.4	94.5	456	AAV24638	H. pylori
9	23.2	80.0	4134	AAx13478	Enterococ
10	23.2	80.0	4134	ABs99273	Enterococ
11	22.4	77.2	110000	13	Continuation (4 of
12	22.2	76.6	4746	10	ADs5747 Escherich
13	22.2	76.6	96589	9	ADA02708 Human ZFH
14	22.2	76.6	96589	10	ADs5747 Escherich
15	22.2	76.6	96589	10	ADs5747 Escherich
16	21.6	74.5	3738	4	ABL01852 Drosophil
17	21.6	74.5	4058	4	ABL03757 Drosophil
18	21.6	74.5	6746	4	ABL03756 Drosophil
19	21.6	74.5	110000	4	Continuation (2 of
20	21.6	74.5	110000	4	Continuation (2 of

21	21.6	74.5	110000	6	ABT00010_01	Continuation (2 of
22	21.6	74.5	110000	6	ABT01503_01	Continuation (2 of
23	21.6	74.5	110000	12	ADH77486_01	ABx67398 Novel Hel
24	21.4	73.8	30	6	ABx67398	Adq97703 Human can
25	21.2	73.1	103665	12	ADQ97703	Aag66950 Cell deat
26	20.6	71.0	3900	2	AAQ66950	Abi08913 Drosophil
27	20.6	71.0	3988	4	ABL08913	ACA33304 Prokaryot
28	20.6	71.0	4278	8	ACA33304	Adh84843 Enterococ
29	20.6	71.0	4320	10	ADH84843	Aax13087 Enterococ
30	20.6	71.0	11739	2	AAx13087	ABs98882 Enterococ
31	20.6	71.0	11739	6	ABs98882	Abi08912 Drosophil
32	20.6	71.0	19867	4	ABL08912	Ades6352 Human PTP
33	20.4	70.3	300000	10	ADs6352	Adci14076 Human mam
34	20.4	70.3	300001	12	ADci14076	ABg83401 Human can
35	20.2	69.7	1863	6	ABQ83401	Adq59380 Human can
36	20.2	69.7	299598	12	ADQ59380	ABk89296 Human gon
37	20.2	69.7	326014	6	ABK89296	Adq94981 Human kin
38	20.2	69.7	326014	12	ADQ94981	Adc93953 E. faeciu
39	20	69.0	876	10	ADC93953	Adc63753 Transcrip
40	20	69.0	1096	12	ADc63753	Abt08124 Myb-relat
41	20	69.0	1192	6	ABT08124	Adh50123 Soybean M
42	20	69.0	1192	12	ADH50123	ACn69837 C. neofo
43	20	69.0	1635	10	ADb69837	ACn89546 Breaet ca
44	20	69.0	1780	11	ACN89546	Adb69476 C. neofo
45	20	69.0	1992	10	ADb69476	

ALIGNMENTS

RESULT 1

AAV07969	AAV07969 standard; DNA; 29 BP.
ID	AAV07969 standard; DNA; 29 BP.
AC	AAV07969;
XX	
DT	25-MAR-2003 (revised)
DT	02-FEB-1999 (first entry)
XX	
DE	Helicobacter pylori polypeptide GHPO 1360 5' DNA primer.
XX	
KW	GHPO 1360; infection; gastritis; ulcer; vaccine; diagnosis; therapy; PCR;
KW	primer; ss.
XX	
OS	Synthetic.
OS	Helicobacter pylori.
XX	
FN	WO9843479-A1.
XX	
PD	08-OCT-1998.
XX	
PF	31-MAR-1998; 98WO-US006421.
XX	
PR	01-APR-1997; 97US-00831310.
XX	
PR	01-APR-1997; 97US-00834666.
XX	
PA	(INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX	(HUMA-) HUMAN GENOME SCI INC.
PI	Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;
XX	
DR	WPI; 1998-568251/48.
XX	
PT	New isolated Helicobacter polynucleotides - used to develop products for
PT	the diagnosis, prevention and treatment of Helicobacter infections and
PT	gastrointestinal diseases.
XX	
PS	Claim 5; Page 155; 184pp; English.
XX	
CC	This 5' primer is used with a 3' primer (see AAV07970) in the PCR
CC	amplification of Helicobacter, e.g. Helicobacter pylori, genomic DNA in
CC	order to obtain DNA (see AAV07963) encoding the unprocessed form of a 32
CC	kDa polypeptide (see AAV73034) designated GHPO 1360. The isolated

CC polynucleotide, and encoded polypeptide, can be used to develop vaccines
CC for the treatment and prevention of Helicobacter infections. (Updated on
CC 25-MAR-2003 to correct PI field.)
XX

SQ Sequence 29 BP; 16 A; 4 C; 5 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 29; DB 2; Length 29;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||
DB 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||

RESULT 2

AAAX14041
ID AAX14041 standard; DNA; 1000 BP.

XX AAX14041;

DT 31-MAR-1999 (first entry)

DE H. pylori GHPO 1360 gene.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.

OS Helicobacter pylori.

FH Key Location/Qualifiers
FT CDS 51..950
/*tag= a

XX WO9843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; 98WO-US006371.

PR 01-APR-1997; 97US-00833457.

PR 24-JUN-1997; 97US-00881227.

PR 29-JUL-1997; 97US-00902615.

XX (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;

XX WPI; 1998-542293/46.

DR P-PSDB; AAW98322.

XX New isolated Helicobacter polynucleotides - used to develop products for

PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastrointestinal diseases.
XX

PS Claim 1; Page 431-432; 2054pp; English.

XX This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these infections,
CC including acute, chronic, and atrophic gastritis, and peptic ulcer
CC diseases, e.g. gastric and duodenal ulcers. They can also be used for the
CC production of antibodies. The products can also be used for detection and
CC diagnosis
XX

SQ Sequence 1000 BP; 372 A; 167 C; 207 G; 254 T; 0 U; 0 Other;

Query Match 100.0%; Score 29; DB 2; Length 1000;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||
DB 37 GTGGAGAACACACAAATGAAAAAATATC 65
|||||

RESULT 3

AAV90844

ID AAV90844 standard; DNA; 1082 BP.

XX AAV90844;

DT 20-MAR-2003 (revised)

DT 18-FEB-1999 (first entry)

DE Nucleotide sequence of cluster 3.

XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.

OS Helicobacter pylori.

XX WO9849314-A2.

PD 05-NOV-1998.

PF 25-APR-1998; 98WO-US008487.

PR 25-APR-1997; 97US-0045107P.

PR 14-OCT-1997; 97US-0061958P.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Chow TP, Fry KE, Lim MY, Mcatee CP;

XX WPI; 1999-009433/01.

XX New Helicobacter pylori antigens and related nucleic acid sequences -
PT useful in serological diagnosis and protective vaccines, providing long-
PT lasting immune response.

XX Claim 20; Page 281-282; 402pp; English.

XX The present sequence encodes a Helicobacter pylori antigenic protein that
CC is characterised by immunoreactivity with H. pylori-positive antisera.
CC The proteins are highly immunogenic and induce a long-lasting immune
CC response that persists even after antimicrobial treatment. In antibody-
CC detection assays, on sera, plasma, urine, saliva etc., they are highly
CC sensitive and specific. The specification also describes 69 previously
CC unrecognised immunogenic cluster families. H. pylori antigens are used to
CC detect H. pylori-specific antibodies, for diagnosing infection or to
CC confirm eradication of infection, and in vaccines to protect against H.
CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
XX

SQ Sequence 1082 BP; 383 A; 188 C; 222 G; 286 T; 0 U; 3 Other;

Query Match 100.0%; Score 29; DB 2; Length 1082;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||
DB 151 GTGGAGAACACACAAATGAAAAAATATC 179
|||||

RESULT 4

AAV90591

ID AAV90591 standard; DNA; 1082 BP.

XX AAV90591;

DT 20-MAR-2003 (revised)

DT 18-FEB-1999 (first entry)

```

XX DE Nucleotide sequence of clone Y104.1.ASM from cluster 3.
XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX OS Helicobacter pylori.
XX PN WO9849314-A2.
XX PD 05-NOV-1998.
XX PF 25-APR-1998; 98WO-US008487.
XX PR 25-APR-1997; 97US-0045107P.
XX PR 14-OCT-1997; 97US-0061958P.
XX PA (GENE-) GENELABS TECHNOLOGIES INC.
XX PI Chow TP, Fry KE, Lim MY, Mcatee CP;
XX WI; 1999-009433/01.
XX PT New Helicobacter pylori antigens and related nucleic acid sequences -
XX PT useful in serological diagnosis and protective vaccines, providing long-
XX PT lasting immune response.
XX PS Claim 27; Page 123-124; 402pp; English.
XX CC The present sequence encodes a Helicobacter pylori antigenic protein that
XX CC is characterised by immunoreactivity with H. pylori-positive antisera.
XX CC The proteins are highly immunogenic and induce a long-lasting immune
XX CC response that persists even after antimicrobial treatment. In antibody-
XX CC detection assays, on sera, plasma, urine, saliva etc., they are highly
XX CC sensitive and specific. The specification also describes 69 previously
XX CC unrecognised immunogenic cluster families. H. pylori antigens are used to
XX CC detect H. pylori-specific antibodies, for diagnosing infection or to
XX CC confirm eradication of infection, and in vaccines to protect against H.
XX CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
XX CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
XX SQ Sequence 1082 BP; 383 A; 189 C; 222 G; 288 T; 0 U; 0 Other;

Query Match 100.0%; Score 29; DB 2; Length 1082;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACATGAAAAAATATC 29
Db 151 GTGGAGAACACACATGAAAAAATATC 179

RESULT 5
AAV90580
ID AAV90580 standard; DNA; 1117 BP.
XX AC AAV90580;
XX DT 20-MAR-2003 (revised)
XX DT 18-FEB-1999 (first entry)
XX DE Nucleotide sequence from clone Y104-1.asm.
XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX OS Helicobacter pylori.
XX PN WO9849314-A2.
XX PD 05-NOV-1998.
XX PF 25-APR-1998; 98WO-US008487.

```

```

XX PR 25-APR-1997; 97US-0045107P.
XX PR 14-OCT-1997; 97US-0061958P.
XX PA (GENE-) GENELABS TECHNOLOGIES INC.
XX PI Chow TP, Fry KE, Lim MY, Mcatee CP;
XX WI; 1999-009433/01.
XX PR P-PSDB; AAW89829.
XX PT New Helicobacter pylori antigens and related nucleic acid sequences -
XX PT useful in serological diagnosis and protective vaccines, providing long-
XX PT lasting immune response.
XX PS Claim 27; Page 119; 402pp; English.
XX CC The present sequence encodes a Helicobacter pylori antigenic protein that
XX CC is characterised by immunoreactivity with H. pylori-positive antisera.
XX CC The proteins are highly immunogenic and induce a long-lasting immune
XX CC response that persists even after antimicrobial treatment. In antibody-
XX CC detection assays, on sera, plasma, urine, saliva etc., they are highly
XX CC sensitive and specific. The specification also describes 69 previously
XX CC unrecognised immunogenic cluster families. H. pylori antigens are used to
XX CC detect H. pylori-specific antibodies, for diagnosing infection or to
XX CC confirm eradication of infection, and in vaccines to protect against H.
XX CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
XX CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
XX SQ Sequence 1117 BP; 388 A; 202 C; 234 G; 293 T; 0 U; 0 Other;

Query Match 100.0%; Score 29; DB 2; Length 1117;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACATGAAAAAATATC 29
Db 169 GTGGAGAACACACATGAAAAAATATC 197

RESULT 6
AAV07963
ID AAV07963 standard; DNA; 1149 BP.
XX AC AAV07963;
XX DT 25-MAR-2003 (revised)
XX DT 02-FEB-1999 (first entry)
XX DE Helicobacter pylori 32 kDa polypeptide GHPO 1360 DNA.
XX KW GHPO 1360; infection; gastritis; ulcer; vaccine; diagnosis; therapy; ss.
XX OS Helicobacter pylori.
XX FH Key Location/Qualifiers
XX CDS 106..1005
XX FT /*tag= a
XX FT sig_peptide 106..166
XX FT /*tag= b
XX FT mat_peptide 167..1002
XX FT /*tag= c
XX PN WO9843479-A1.
XX PD 08-OCT-1998.
XX PF 31-MAR-1998; 98WO-US006421.
XX PR 01-APR-1997; 97US-00831310.
XX PR 01-APR-1997; 97US-00834666.
XX PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

```

```

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kleanthous H, Lissolo L, Tomb J, Miller C, Algarawi A;
XX
DR WPI; 1998-568251/48.
XX P-PSDB; AAW73034.
XX
PT New isolated Helicobacter polynucleotides - used to develop products for
PT the diagnosis, prevention and treatment of Helicobacter infections and
PT gastroduodenal diseases.
XX
PS Claim 1; Page 148-149; 184pp; English.
XX
CC This DNA sequence codes for a 32 kDa Helicobacter pylori polypeptide (see
CC AAW73034) designated GHPO 1360. A polynucleotide encoding the
CC matured, or the mature, GHPO 1360 can be obtained from genomic DNA by
CC PCR amplification (see AAV07969-70 and AAV07973). The invention provides
CC polynucleotides (see AAV72001, AAV07912-21 and AAV07963-64) encoding a
CC family of 76 kDa Helicobacter polypeptides (see AAW73022-32), GHPO 1360
CC and a 50 kDa polypeptide (see AAW73035). These polynucleotides were
CC initially identified in a search of H. pylori genomic databases. DNA
CC cassettes for expression of the Helicobacter proteins (unprocessed or
CC mature forms) in prokaryotic or eukaryotic cells are provided. The
CC polynucleotides can be used in vaccines to prevent or treat Hb infection
CC in a mammal. Viral (especially poxvirus) or bacterial vectors are used.
CC Products and methods of the invention allow treatment and prevention of
CC gastroduodenal diseases associated with Hb infections, including acute,
CC chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric
CC and duodenal ulcers. Diagnostic and detection methods are also provided.
CC GHPO 1360 was demonstrated to be a protective antigen. (Updated on 25-MAR
CC -2003 to correct PI field.)
XX
SQ Sequence 1149 BP; 412 A; 185 C; 242 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 29; DB 2; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACAAATGAAAAAATATC 29
DB 92 GTGGAGAACACACAAATGAAAAAATATC 120

RESULT 7
AAD61576
ID AAD61576 standard; DNA; 1149 BP.
XX
AC AAD61576;
XX
DT 15-JAN-2004 (first entry)
DE Helicobacter pylori p32 antigen DNA.
XX
KW Immune response; Helicobacter infection; adjuvant; heat-labile toxin; LT;
KW cholera toxin; CT; urease; therapy; antibacterial; vaccine; antigen; p32;
KW gene; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 106..1005
FT /tag= a
FT /product= "Antigen p32"
FT sig_peptide 106..166
FT /tag= b
FT mat_peptide 167..1002
FT /tag= c
FT /product= "Mature p32"
XX
XX US676244-B1.
XX
XX 10-JUN-2003.
XX

PF 18-JUN-1999; 99US-00336115.
XX
PR 19-JUN-1998; 98US-00100258.
XX (ACAM-) ACAMBIS INC.
XX
PI Weltzin RA, Guy B;
XX
XX WPI; 2003-799824/75.
DR P-PSDB; ABW00787.
XX
PT Inducing immune response to Helicobacter useful for treating Helicobacter
PT pylori infection, by administering immunogenic Helicobacter polypeptide
PT admixed with adjuvant having heat-labile toxin of Escherichia coli.
XX
PS Claim 8; Col 105-108; Opp; English.
XX
CC The invention relates to a method for inducing protective immune response
CC to Helicobacter infection. The method comprising administering to a
CC mammal by injection an immunogenic Helicobacter pylori polypeptide
CC comprising a subunit of H. pylori urease admixed with an adjuvant having
CC one or more heat-labile toxin of Escherichia coli (LT), B subunit of LT
CC (LTB), cholera toxin (CT), and B subunit of CT. The method is useful for
CC inducing an immune response to Helicobacter infection in a mammal. The
CC method is useful for both treatment and prevention of H. pylori infection.
CC The invention is useful as vaccine. The present sequence is Helicobacter
CC pylori p32 antigen DNA
XX
SQ Sequence 1149 BP; 412 A; 185 C; 242 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 29; DB 10; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACAAATGAAAAAATATC 29
DB 92 GTGGAGAACACACAAATGAAAAAATATC 120

RESULT 8
AAV24638
ID AAV24638 standard; DNA; 456 BP.
XX
AC AAV24638;
XX
DT 26-JUN-1998 (first entry)
DE H. pylori ORF 03xell1215orf7.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW bacterium; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1..456
FT /tag= a
FT /note= "no stop codon given"
XX
XX WO9737044-A1.
XX
PD 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US005223.
XX
XX 29-MAR-1996; 96US-00625811.
XX 02-APR-1996; 96US-00758731.
XX 25-OCT-1996; 96US-00736905.
XX 28-OCT-1996; 96US-00738859.
XX 06-DEC-1996; 96US-00761318.
XX

```

PA (ASTR) ASTRA AB.
 XX Smith D, Alm RA;
 XX WPI: 1997-503122/46.
 DR P-PSDB; ANW55229.
 XX Helicobacter pylori nucleic acid sequences and encoded polypeptide(s) -
 PT useful in vaccines to treat or prevent H. pylori infection and for
 PT diagnosis of H. pylori infection.
 XX Claim 5,6; Page 155; 1145pp; English.
 XX This sequence encodes a Helicobacter pylori protein of unspecified
 CC function. The protein may be used in a vaccine to prevent or treat H.
 CC pylori infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of H.
 CC pylori in a sample, and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences, and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts
 XX
 SQ Sequence 456 BP; 170 A; 79 C; 106 G; 101 T; 0 U; 0 Other;
 Query Match 94.5%; Score 27.4; DB 2; Length 456;
 Best Local Similarity 96.6%; Pred. No. 2.4;
 Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTGGAGACACACAAATGAAAAAATATC 29
 DB 2 GTGGAGACACACAAATGAAAAAATATC 30
 RESULT 9
 AAX13478/C
 ID AAX13478 standard; DNA; 4134 BP.
 XX
 AC AAX13478;
 XX
 DT 19-MAR-1999 (first entry)
 XX
 DE Enterococcus faecalis genome contig SEQ ID NO:541.
 XX
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 XX
 OS Enterococcus faecalis.
 XX
 FN WO980555-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US008985.
 XX
 PR 06-MAY-1997; 97US-0044031P.
 PR 16-MAY-1997; 97US-0046655P.
 PR 14-NOV-1997; 97US-0066009P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 FI Kunsch CA, Dillon PJ, Barash SC;
 WPI: 1999-045171/04.
 New isolated Enterococcus faecalis polynucleotides and polypeptides -
 used to develop products for the detection of Enterococcus and for use in
 vaccines for prevention or attenuation of Enterococcus infection.
 Claim 1; Page 1795-1797; 2084pp; English.
 A computer readable medium has been developed which has recorded on it
 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 AAX12938 to AAX13919 represent these nucleotide sequences which are
 primary nucleotide sequences, also known as contigs. The computer-based
 system can identify fragments of the Enterococcus faecalis genome with
 commercial importance. The products can be used to detect the presence of
 Enterococcus faecalis in samples. They can also be used for diagnosing
 CC Enterococcal infection in an animal and monitoring progression of
 CC disease, and for identifying agents which can be used to modulate the
 CC growth or pathogenicity of Enterococcus faecalis, or another related
 CC organism, in vivo or in vitro. In particular the polypeptides encoded by
 CC the Enterococcus faecalis nucleotide sequences can be used in vaccines to
 CC prevent or attenuate an Enterococcal infection
 XX
 SQ Sequence 4134 BP; 1370 A; 863 C; 578 G; 1315 T; 0 U; 8 Other;
 Query Match 80.0%; Score 23.2; DB 2; Length 4134;
 Best Local Similarity 89.3%; Pred. No. 81;
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TGGAGACACACAAATGAAAAAATATC 29
 DB 1851 TGGAGACACACAAATGAAAAAATATC 1824
 RESULT 10
 ABS99273/C
 ID ABS99273 standard; DNA; 4134 BP.
 XX
 AC ABS99273;
 XX
 DT 18-DEC-2002 (first entry)
 XX
 DE Enterococcus faecalis contig sequence #541.
 XX
 KW Computer readable medium; Enterococcus faecalis; microbe; growth;
 KW pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
 KW therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
 KW biotech technology; antibacterial; modulator of nucleic acid expression;
 KW contig; ds.
 XX
 OS Enterococcus faecalis.
 XX
 FN US2002120116-A1.
 XX
 PD 29-AUG-2002.
 XX
 PF 04-MAY-1998; 98US-00070927.
 XX
 PR 04-MAY-1998; 98US-00070927.
 XX
 PA (KUNSCH) KUNSCH C A.
 PA (DILLON) DILLON P J.
 PA (BARASH) BARASH S.
 XX
 PI Kunsch CA, Dillon PJ, Barash S;
 XX
 DR WPI: 2002-750065/81.
 XX
 PT Computer readable medium having recorded on it a Enterococcus faecalis
 PT nucleotide sequence useful for detecting diseases related to Enterococcus
 PT infections in animals.
 XX
 PS Claim 1; Page; 119pp; English.

XX The present invention relates to a new computer readable medium with an
CC Enterococcus faecalis nucleotide sequence. The invention is useful to
CC diagnose the presence of E.faecalis in a sample or determining the
CC presence of a specific microbe in a sample. The invention is also useful
CC for modulating the growth or pathogenicity of E.faecalis, in a vaccine to
CC confer resistance to Enterococcal infection, for commercial, therapeutic
CC and industrial purposes, and for fermenting a particular sugar source or
CC to produce a particular metabolite. The invention is useful for detecting
CC diseases related to Enterococcus infections in animals, and for detecting
CC E.faecalis using biotech technology. The present nucleic acid sequence
CC represents an Enterococcus faecalis contig DNA sequence of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at http.sequedata.uspto.gov
XX

PN Sequence 4134 BP; 1370 A; 863 C; 578 G; 1315 T; 0 U; 8 Other;

Query Match 80.0%; Score 23.2; DB 6; Length 4134;
Best Local Similarity 89.3%; Pred. No. 81;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TGGAGACACACAAATGAAAAAATATC 29

Db 1851 TGGAGACACACAAATGAAAAAATATC 1824

RESULT 11

ABD32627_3
Continuation (4 of 5) of ABD32627 from base 300001 (Mouse cancer-associated genomic DNA
WP Sequence split into 5 fragments LOCUS ABD32627 Accession ABD32627

WP	Fragment Name	Begin	End
WP	ABD32627_0	1	110000
WP	ABD32627_1	100001	210000
WP	ABD32627_2	200001	310000
WP	ABD32627_3	300001	410000
WP	ABD32627_4	400001	421609

Query Match 77.2%; Score 22.4; DB 13; Length 110000;
Best Local Similarity 95.8%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 GAACACACATGAAAAAATATC 29

Db 19593 GAACACACATGAAAAAATATC 19616

RESULT 12

AAD55747
ID AAD55747 standard; DNA; 4746 BP.

XX AAD55747;

DT 07-AUG-2003 (first entry)

DE Escherichia coli CS3 operon.

XX Coli surface antigen; CS; vaccine; diarrhoea; bacterial infection;
KW antigen; cstB; cstA; cstC; cstE; cstF; cstG; cstH; gene; ds.

OS Escherichia coli.

FT	Key	Location/Qualifiers
FT	misc_feature	378..1103
FT		/tag= a
FT		/note= "cstA gene"

FT	misc_feature	1858..3579
FT		/tag= b
FT		/note= "cstB gene"

FT	misc_feature	2266..3579
FT		/tag= c
FT		/note= "cstC gene"

FT	misc_feature	2668..3579
----	--------------	------------

FT		/tag= d
FT	misc_feature	3031..3579
FT		/tag= e
FT		/note= "cstE gene"
FT	misc_feature	3577..3579
FT		/tag= f
FT		/note= "cstF gene"
FT	misc_feature	3580..4131
FT		/tag= g
FT		/note= "cstG gene"
FT	misc_feature	4153..4659
FT		/tag= h
FT		/note= "cstH gene"

PN WO2003022306-A2.

XX PD 20-MAR-2003.

XX PF 11-SEP-2002; 2002WO-GB004123.

XX PR 11-SEP-2001; 2001GB-00021998.

XX PA (ACAM-) ACAMBIS RES LTD.

XX PI Turner AK, Greenwood J, Stephens JC, Beavis JC, Darsley MJ;

XX DR WPI; 2003-301009/29.

XX New bacterial cell expressing three or more coli surface antigens, useful
PT for manufacturing a medicament, i.e. a vaccine, for vaccination against
PT diarrhea.

XX PS Disclosure; Page 102-104; 58pp; English.

XX The invention relates to a bacterial cell expressing three or more coli
CC surface (CS) antigens. The bacterial cell is useful for manufacturing a
CC medicament, i.e. a vaccine, for vaccination against diarrhoea. The
CC vaccine is also useful for targeting bacterial infection. The present
CC sequence is Escherichia coli CS3 operon containing cstB, cstA, cstC, cstD,
CC cstE, cstF, cstG and cstH gene. This sequence is used in the invention

SQ Sequence 4746 BP; 1588 A; 745 C; 957 G; 1456 T; 0 U; 0 Other;

Query Match 76.6%; Score 22.2; DB 10; Length 4746;

Best Local Similarity 88.9%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TGGAGACACACAAATGAAAAAATAT 28

Db 3631 TGGAGACACACACAGTGAATAATCAT 3657

RESULT 13

ADA02708/c

ID ADA02708 standard; DNA; 96589 BP.

XX ADA02708;

XX 06-NOV-2003 (first entry)

XX Human ZFX1B carcinoma associated gene, SEQ ID NO:1226.

XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.

OS Homo sapiens.

XX WO2003057146-A2.

XX PD 17-JUL-2003.

XX

PF 26-DEC-2002; 2002WO-US041414.
XX 26-DEC-2001; 2001US-00035832.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX WPI; 2003-587068/55.
XX New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.
XX Claim 1; SEQ ID NO 1226; 245pp; English.
XX The invention relates to recombinant carcinoma associated (CA) nucleic
XX acid sequences from mouse and human (ADA01482-ADA03094), and to
XX recombinant carcinoma associated proteins (CAP) encoded by them. The
XX invention also encompasses expression vectors and host cells comprising a
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically
XX binds to the protein, and a biochip comprising CA nucleic acid or
XX fragments thereof. The sequences of the invention were identified using
XX oncogenic retroviruses, which insert into the genome of the host organism
XX at random. Many of these do not carry transduced host oncogenes or
XX pathogenic trans-acting viral genes, meaning that cancer incidence is a
XX direct consequence of the effects of proviral integration into host
XX protooncogenes. The CA nucleic acid sequences can be used to diagnose
XX carcinoma (especially breast cancer, prostate cancer, lymphoma or
XX leukaemia) or a propensity to carcinoma by determination of the sequence
XX of a CA gene, or by determination of CA gene expression in particular
XX tissues. CA nucleic acids, proteins and antibodies are also useful as
XX therapeutic agents and in screening and evaluating drug candidates. The
XX present sequence represents a specifically claimed human CA nucleic acid
XX sequence of the invention. Note: The complete sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 96589 BP; 29004 A; 18509 C; 19494 G; 29562 T; 0 U; 20 Other;
Query Match 76.6%; Score 22.2; DB 9; Length 96589;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 GGAGAACACACAAATGAAAAAATATC 29
Db 40453 GCAGACACAAAGGAAAAAATATC 40427

RESULT 14
ADB72446/c
ID ADB72446 standard; DNA; 96589 BP.
XX
XX ADB72446;
XX
XX 04-DEC-2003 (first entry)
XX Human ZFX1B gene.
XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX Homo sapiens.
XX WO2003008583-A2.
XX
XX 30-JAN-2003.
XX
XX 26-DEC-2001; 2001WO-US051291.
XX
XX 02-MAR-2001; 2001US-00798586.
XX 23-OCT-2001; 2001US-00004113.
XX 08-NOV-2001; 2001US-00052482.

PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW, Engelhard EK;
XX WPI; 2003-239337/23.
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX cancers, neoplasm, adenocarcinoma, or sarcomas.
XX Claim 1; SEQ ID NO 274; 2304pp; English.
XX The invention relates to a novel recombinant nucleic acid comprising a
XX nucleotide sequence selected from any of the 660 sequences fully defined
XX in the specification. A polynucleotide of the invention has cytostatic
XX activity, and may have a use in gene therapy, or in a vaccine. The
XX recombinant nucleic acids and polypeptides are useful for treating
XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX sarcomas. The present sequence represents a human gene of the invention.
XX Sequence 96589 BP; 29004 A; 18509 C; 19494 G; 29562 T; 0 U; 20 Other;
Query Match 76.6%; Score 22.2; DB 10; Length 96589;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 GGAGAACACACAAATGAAAAAATATC 29
Db 40453 GCAGACACAAAGGAAAAAATATC 40427

RESULT 15
ADE95956/c
ID ADE95956 standard; DNA; 96589 BP.
XX
XX ADE95956;
XX
XX 12-FEB-2004 (first entry)
XX Human ZFX1B gene genomic DNA sequence.
XX cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
XX lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; ZFX1B.
XX Homo sapiens.
XX WO2003039484-A2.
XX
XX 15-MAY-2003.
XX
XX 08-NOV-2002; 2002WO-US036071.
XX
XX 08-NOV-2001; 2001US-00052482.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW, Engelhard EK;
XX WPI; 2003-441462/41.
XX New carcinoma associated nucleic acids and proteins, useful for screening
XX drug candidates, or for diagnosing and treating carcinomas, e.g.
XX lymphoma, breast cancer, prostate cancer or leukemia.
XX Claim 1; SEQ ID NO 214; 793pp; English.
XX This invention relates to novel recombinant nucleic acids for use in
XX diagnosis and treatment of cancer, especially carcinomas, as well as the
XX use of compositions in screening methods. The compositions of the
XX invention may have cytostatic activity whilst the disclosed sequences may
XX be useful for gene therapy. The carcinoma associated nucleic acids and

CC proteins are useful for diagnosing and treating carcinomas, for example
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC drug candidates or bioactive agents capable of binding to, or modulating
CC the activity of, a carcinoma associated protein. The present sequence is
CC the genomic DNA sequence of the human ZFX1B gene which is a carcinoma
CC associated gene of the invention.

XX

SQ Sequence 96589 BP; 29004 A; 18508 C; 19495 G; 29562 T; 0 U; 20 Other;

Query Match 76.6%; Score 22.2; DB 10; Length 96589;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GGAGAACACACAAATGAAAAAATATC 29
Db 40453 GCAGAACACAAAGGAAAAAATATC 40427

Search completed: April 24, 2005, 02:20:02
Job time : 194.542 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 23, 2005, 17:27:10 ; Search time 752.775 Seconds
(without alignments)
1866.696 Million cell updates/sec

Title: US-10-039-183A-14
Perfect score: 29
Sequence: 1 gtggagacacacaaatgaaaaaaatc 29

Scoring table: IDENTITY_NUC
Gapop 10.0' , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	29	6	BD082352
2	29	100.0	1000	6	BD092341
3	29	100.0	1082	6	BD061716
4	29	100.0	1082	6	BD061969
5	29	100.0	1117	6	BD061705
6	29	100.0	1149	6	AR342398
7	29	100.0	1149	6	BD082346
8	29	100.0	10085	1	AE001454
9	29	100.0	11421	1	AE000538
10	23.2	80.0	4134	6	BD193947
11	23.2	80.0	304454	1	AE016956
12	22.6	77.9	110098	9	AL358976
13	22.6	77.9	171692	5	BX088597
14	22.6	77.9	173670	9	AL133351
15	22.6	77.9	189058	2	AC013339
16	22.4	77.2	113764	9	HS1651E10
17	22.4	77.2	121609	6	CQ870465
18	22.4	77.2	199443	2	AL354988
19	22.4	77.2	349980	6	CQ869727

20	22.2	76.6	4746	1	ECCS3P
21	22.2	76.6	4746	6	AX741417
22	22.2	76.6	94252	8	AC137616
23	22.2	76.6	96589	6	AX695599
24	22.2	76.6	126080	9	AY029472
25	22.2	76.6	126332	8	AC137621
26	22.2	76.6	159791	9	AC009951
27	22.2	76.6	230342	2	AC106104
28	22.2	76.6	286655	2	AC098489
29	21.8	75.2	184455	2	AC148367
30	21.8	75.2	193506	10	AC116953
31	21.8	75.2	217538	2	AC098034
32	21.6	74.5	3738	6	CQ572261
33	21.6	74.5	4058	6	CQ575118
34	21.6	74.5	4173	3	AY061475
35	21.6	74.5	6746	6	CQ575117
36	21.6	74.5	59411	9	AC004040
37	21.6	74.5	67847	2	AC017241
38	21.6	74.5	95539	3	AC084471
39	21.6	74.5	96924	5	EX890635
40	21.6	74.5	103034	2	AC067978
41	21.6	74.5	110000	9	AF491780_01
42	21.6	74.5	111934	9	AF281074
43	21.6	74.5	116039	9	AC007362
44	21.6	74.5	150229	5	EX001056
45	21.6	74.5	156855	9	AC104000

ALIGNMENTS

RESULT 1
BD082352
LOCUS
DEFINITION
29 bp DNA linear PAT 27-AUG-2002
76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and corresponding polynucleotide molecules.
ACCESSION
BD082352
VERSION
BD082352.1 GI:22627962
KEYWORDS
JP 2001523954-A/59.
SOURCE
Mastadenovirus
ORGANISM
Mastadenovirus
REFERENCE
1 (bases 1 to 29)
AUTHORS
Kleanthous, H., Lissolo, L., Tomb, J.F., Miller, C. and Garawi, A.A.
TITLE
76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and corresponding polynucleotide molecules
JOURNAL
Patent: JP 2001523954-A 59 27-NOV-2001;
MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
HUMAN GENOME SCIENCES INC
COMMENT
PN JP 2001523954-A/59
PD 27-NOV-2001
PF 31-MAR-1998 JP 1998541962
PI HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI MILLER,
ANAL AL GARAWI
PI A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,
PC A61K39/40,
PC G01N33/554, G01N33/569
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
1..29
/organism="Mastadenovirus"
/mol_type="genomic DNA"
/db_xref="taxon:10509"

ORIGIN

Query Match 100.0%; Score 29; DB 6; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||
Db 1 GTGGAGAACACACAAATGAAAAAATATC 29

RESULT 2
LOCUS BD092341 1000 bp DNA linear PAT 27-AUG-2002
DEFINITION Identification of polynucleotides encoding novel helicobacter
ACCESSION BD092341
VERSION BD092341.1 GI:22637952
KEYWORDS JP 2001527393-A/122.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1000)
AUTHORS Kleanthous, H., Garawi, A.A., Miller, C., Tomb, J.F. and Oomen, R.P.
TITLE Identification of polynucleotides encoding novel helicobacter
JOURNAL polypeptides in the helicobacter genome
Patent: JP 2001527393-A 122 25-DEC-2001;
MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET
VACCINS AGROBIOLOGICAL RESOURCES MINISTRY O SA, HUMAN GENOME
SCIENCES INC
COMMENT PN JP 2001527393-A/122
PD 25-DEC-2001
PF 01-APR-1998 JP 1998541947
PR 01-APR-1997 US 08/833457, 24-JUN-1997 US 08/881227 PR
29-JUL-1997 US 08/902615
PI HAROLD KLEANTHOUS, AMAL AL GARAWI, CHARLES MILLER, JEAN FRANCOIS
TOMB,
PI RAYMOND PETER OOMEN
PC A01N43/04, A61K31/70
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source 1..1000
/mol_type="synthetic construct"
/db_xref="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 100.0%; Score 29; DB 6; Length 1000;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||
Db 37 GTGGAGAACACACAAATGAAAAAATATC 65

RESULT 3
LOCUS BD061716 1082 bp DNA linear PAT 27-AUG-2002
DEFINITION Antigenic composition and method of detection for Helicobacter
ACCESSION BD061716
VERSION BD061716.1 GI:22607321
KEYWORDS JP 2001517091-A/50.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1082)
AUTHORS Chow, T.P., Fry, K.E., Lim, M.Y. and Mcatee, C.P.
TITLE Antigenic composition and method of detection for Helicobacter
JOURNAL Patent: JP 2001517091-A 50 02-OCT-2001;
GENELABS TECHNOLOGIES INC
COMMENT PN JP 2001517091-A/50
PD 02-OCT-2001
PF 25-APR-1998 JP 1998547263
PR 25-APR-1997 US 60/045107, 14-OCT-1997 US 60/061958 PI
THERESA P CHOW, KIRK E FRY, MOON Y LIM, C P MCATEE PC
Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||
Db 37 GTGGAGAACACACAAATGAAAAAATATC 65

RESULT 4
LOCUS BD061969 1082 bp DNA linear PAT 27-AUG-2002
DEFINITION Antigenic composition and method of detection for Helicobacter
ACCESSION BD061969
VERSION BD061969.1 GI:22607574
KEYWORDS JP 2001517091-A/303.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1082)
AUTHORS Chow, T.P., Fry, K.E., Lim, M.Y. and Mcatee, C.P.
TITLE Antigenic composition and method of detection for Helicobacter
JOURNAL Patent: JP 2001517091-A 303 02-OCT-2001;
GENELABS TECHNOLOGIES INC
COMMENT PN JP 2001517091-A/303
PD 02-OCT-2001
PF 25-APR-1998 JP 1998547263
PR 25-APR-1997 US 60/045107, 14-OCT-1997 US 60/061958 PI
THERESA P CHOW, KIRK E FRY, MOON Y LIM, C P MCATEE PC
Qy 1 GTGGAGAACACACAAATGAAAAAATATC 29
|||||
Db 151 GTGGAGAACACACAAATGAAAAAATATC 179

RESULT 5
LOCUS BD061705 1117 bp DNA linear PAT 27-AUG-2002
DEFINITION Antigenic composition and method of detection for Helicobacter
ACCESSION BD061705
VERSION BD061705.1 GI:22607310
KEYWORDS JP 2001517091-A/39.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1117)
```

AUTHORS Chow, T.P., Fry, K.E., Lim, M.Y. and Mcatee, C.P.
TITLE Antigenic composition and method of detection for Helicobacter
JOURNAL Patent: JP 2001517091-A 39 02-OCT-2001;
COMMENT GENELABS TECHNOLOGIES INC
 PN JP 2001517091-A/39
 PD 02-OCT-2001
 PF 25-APR-1998 JP 1998547263
 FR 25-APR-1997 US 60/045107,14-OCT-1997 US 60/061958 PI
 THERESA P CHOW, KIRK E FRY, MOON Y LIM, C P MCATEE PC
 C12N15/31, C07K14/205, C07K16/12, A61K39/106
 CC Strandedness: Single;
 CC Topology: Linear;
 FH Key Location/Qualifiers.

FEATURES
 source
 1..117
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"

ORIGIN

Query Match 100.0%; Score 29; DB 6; Length 1117;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGAGAACACACATGAAAAAATATC 29
 |||||
Db 169 GTGGAGAACACACATGAAAAAATATC 197
 |||||

RESULT 6
AR342398
LOCUS AR342398 1149 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 23 from patent US 6576244.
ACCESSION AR342398
VERSION AR342398.1 GI:33737371
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1149)
AUTHORS Weltzin, R.A. and Guv, B.
TITLE LT and CT in parental immunization methods against helicobacter infection
JOURNAL Patent: US 6576244-A 23 10-JUN-2003;
FEATURES Location/Qualifiers
 source
 1..1149
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 29; DB 6; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGAGAACACACATGAAAAAATATC 29
 |||||
Db 92 GTGGAGAACACACATGAAAAAATATC 120
 |||||

RESULT 7
BD082346
LOCUS BD082346 1149 bp DNA linear PAT 27-AUG-2002
DEFINITION 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and corresponding polynucleotide molecules.
ACCESSION BD082346
VERSION BD082346.1 GI:22627956
KEYWORDS JP 2001523954-A/53.
SOURCE Mastadenovirus
ORGANISM Mastadenovirus
REFERENCE 1 (bases 1 to 1149)
AUTHORS Kleanthous, H., Lissolo, L., Tomb, J.F., Miller, C. and Garawi, A.A.
TITLE 76 kDa, 32 kDa, and 50 kDa helicobacter polypeptides and

corresponding polynucleotide molecules
 Patent: JP 2001523954-A 53 27-NOV-2001;
 MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR MERIEUX SERUMS ET
 VACCINS SECRETARY OF THE DEPARTMENT OF HEALTH HUMAN SERVICES SA,
 HUMAN GENOME SCIENCES INC
 PN JP 2001523954-A/53
 PD 27-NOV-2001
 PF 31-MAR-1998 JP 1998541962
 PI HAROLD KLEANTHOUS, LING LISSOLO, JEAN FRANCOIS TOMB, CHARLES PI
 MILLER,
 PI AMAL AL GARAWI
 PC A01N43/04, A01N59/16, A61K9/48, A61K31/70, A61K31/715, A61K39/02,
 PC A61K39/40,
 PC G01N33/54, G01N33/569
 CC Strandedness: Single;
 CC Topology: Linear;
 FH Key Location/Qualifiers.

FEATURES
 source
 1..1149
 /organism="Mastadenovirus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10509"

ORIGIN

Query Match 100.0%; Score 29; DB 6; Length 1149;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGAGAACACACATGAAAAAATATC 29
 |||||
Db 92 GTGGAGAACACACATGAAAAAATATC 120
 |||||

RESULT 8
AE001454
LOCUS AE001454 10085 bp DNA linear BCT 20-JAN-1999
DEFINITION Helicobacter pylori, strain J99 section 15 of the complete genome.
ACCESSION AE001454 AE001439
VERSION AE001454.1 GI:4154666
KEYWORDS
SOURCE Helicobacter pylori J99
ORGANISM Helicobacter pylori J99
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 Helicobacteraceae; Helicobacter.
 1 (bases 1 to 10085)
 Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C.,
 Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G.,
 Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C.,
 Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E.,
 Vovis, G.F. and Trust, T.J.
 Genomic-sequence comparison of two unrelated isolates of the human
 gastric pathogen Helicobacter pylori
 Nature 397 (6715), 176-180 (1999)
 99120557
 9923682
 2 (bases 1 to 10085)
 King, B.L., Alm, R.A. and Trust, T.J.
 Direct Submission
 Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney
 Street, Cambridge, MA 02139, USA
 Address all correspondence to: hp@arc.us.astro.com or Richard
 A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,
 MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,
 Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,
 Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics
 Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and
 Diane E. Taylor are with the University of Alberta Department of
 Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G
 2H7 and the Canadian Bacterial Diseases Network. All other authors
 are with Astra Research Center Boston, 128 Sidney Street,
 Cambridge, MA, 02139. Putative identifications, sequence
 alignments, and name and sequence search capability are available

at ARCB's World Wide Web site. (URL:
http://www.astro-boston.com/hpylori).

FEATURES	source	Location/Qualifiers
gene	1. .10085	/organism="Helicobacter pylori J99"
		/mol_type="genomic DNA"
		/strain="J99"
		/db_xref="taxon:85963"
		complement(235. .708)
CDS		/gene="jhp0153"
		complement(235. .708)
		/gene="jhp0153"
		/note="similar to H. pylori 26695 gene HP0167"
		/codon_start=1
gene		/transl_table=11
		/product="putative"
		/protein_id="AAD05736.1"
		/db_xref="GI:4154669"
		/translation="MAGRMELKNIISETLNIEKMAKTIIDGFNTAQKTPGFFKTPSH LQNTPSKNANTPLPKNAKIEIQEKITEKEEAKELITEIIVQENTQALINPVRV FLKGLERTVLVFKGMQALDEEKAUKLDLVARFIQYQLSVLEKRLSELRNTE"
CDS		complement(687. .950)
		/gene="jhp0154"
		complement(687. .950)
		/gene="jhp0154"
		/note="similar to H. pylori 26695 gene HP0168"
gene		/codon_start=1
		/transl_table=11
		/product="putative"
		/protein_id="AAD05737.1"
		/db_xref="GI:4154670"
CDS		/translation="MPLTITLAIHYEQGFPEERALQIYTNILKKTPTONTALKOMKR LEKIQKCAPFKHDLAKQHYLNFIKGRDLSVENLEKWLVEWN"
		complement(997. .2265)
		/gene="jhp0155"
		complement(997. .2265)
		/gene="jhp0155"
gene		/note="similar to H. pylori 26695 gene HP0169"
		/codon_start=1
		/transl_table=11
		/product="putative PROTEASE"
		/protein_id="AAD05738.1"
CDS		/db_xref="GI:4154671"
		/translation="MNQVELLSAPGNLKKLIALNYGADAVYGVSHFSLNRACKEF TLETKEGIDYAHLDKKVYATINGFPFNSQLKLEELHYKMAELDEPAFIIAPGVI KLAKIAPHI PVHLSQANVLNLTDAQFYDLGVKRIVCARELSLNDAVEIKKALPDL ELEIFVHSGMCFAPSGRCCLISALQGRVFNRGSCANDCRFDYEVYVKNPDNGVMWRLV BEEGVGTHIFNAKDLSHGHTAEILSSNAISALKIEGRTKSSYAAQTRIVRLAVDD FYHNTLKPSFYASELNTLKNRGFTDGYLMWRPFELDTQNHQTAISEGDFVNGEITE DGRFPACFTTTNTAYELIAPKNAAITPINDIGKIYTFEKRSLVLYKILLENNTE LETHSGNVNLVRLPAPLPAPSFRLTQVRV"
		complement(2268. .2999)
		/gene="jhp0156"
		complement(2268. .2999)
gene		/gene="jhp0156"
		/note="similar to H. pylori 26695 gene HP0170"
		/codon_start=1
		/transl_table=11
		/product="putative"
CDS		/protein_id="AAD05739.1"
		/db_xref="GI:4154672"
		/translation="MTQBELDALMSGDLLENLEALEAKEAKEPTKEPKSKENSNY SEKMTVKKEDEAKYKISPHWPPPPPTTEEHKVHQLDDVTRDSEVKATQIFDQDLI GASAEKIAWKVKIQEIPKQIEIFDNLHAHPHVESFTALNEQOEITLNAKISLEE AANCSDSQAMDIMQFQDIHRQKIERVVNVNRALSQYMSNLSIFEKGIIDSKRVSSATY ITGDDDKDLASADDIEALIASFGAK"
		complement(3069. .4160)
		/gene="prfb"
gene		/gene="prfb"
		/note="synonym: jhp0157"
		complement(3069. .4160)
		/gene="prfb"
		/note="similar to H. pylori 26695 gene HP0171"
CDS		/codon_start=1
		/transl_table=11
		/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
		/protein_id="AAD05744.1"
		/db_xref="GI:4154677"
gene		/translation="MKKNILNALVGLSASFLMAKPAHNANNSTHNTKETTDSAGV LATVDGRPTIKQRPDMIKQRPNFDFOKLKEKEKEALTEQAIRTALVNEAKAEKLNQ TPEFKAMWEAVKQALVEFWAKQAEVKKIQIPEKEMQDFYNANKDQLFVKQEAHAR
		complement(4221. .5393)
		/gene="moea"
		/note="synonym: jhp0158"
		complement(4221. .5393)
CDS		/gene="moea"
		/note="similar to H. pylori 26695 gene HP0172"
		/codon_start=1
		/transl_table=11
		/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
gene		/protein_id="AAD05741.1"
		/db_xref="GI:4154674"
		/translation="MISFKEALKITHSNIPKLEPVEVVSLEFESAGRILAEILICVHAL PKFQNSAMDGCFKMDIGQKTQIHI FAGDVSALVENECVKIMTGAMVPKGIE TIVFIECMLESHKDFALAPKDFKIHANIQKGENASNLVLPKNTLNYNTHGIALIAS OGPKEIKAFRLKIALPSSGDELVPAGONALECOVVDYNSGVENMLKNYNTHFTGLV KDDKNLQKILELQGYDVLSSAGVSVDGDFPKDALKERNALFYEKVNLPKGPVT LAQLNQSIILGPGNLSCLLVRLVILFLLESLNLDKFKLPKPAQINAPLKNKK RTHILGNYSNHQFIPYNNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFEN"
		complement(5403. .6170)
		/gene="flir"
CDS		/note="synonym: jhp0159"
		complement(5403. .6170)
		/gene="flir"
		/note="similar to H. pylori 26695 gene HP0173"
		/codon_start=1
gene		/transl_table=11
		/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
		/protein_id="AAD05742.1"
		/db_xref="GI:4154675"
		/translation="MLDPIQELSTPHVRDFLLFLRVSGVLSFPFFFNHVLVPLSVRG ALSLYVAIFYPTLEFNAIYTPGFIACELCFLGVCASVFFQIVFASLVFAFDSI SFSMGLTMASAYDPISGOKPIVGOALLLAILLLDLSFHHOILLFVDHSLKAVPLG RFVPEPELAKNIVKAFSHLVIGFSMAPPIICLVLLSDIIFGMIMKTHPQFNALLAIFG PVKIAIGFVGHILLASAIMGRFKEISLAFSVISKIP"
CDS		complement(6164. .6892)
		/gene="jhp0160"
		complement(6164. .6892)
		/gene="jhp0160"
		/note="similar to H. pylori 26695 gene HP0174"
gene		/codon_start=1
		/transl_table=11
		/product="putative"
		/protein_id="AAD05743.1"
		/db_xref="GI:4154676"
CDS		/translation="MVLFLSIFKSFNDFLSARMLLINLGPILLSLAFSGAIFYVNGG SIVNYCQTLLEQSLNDYAHSGQFFAGVPAWFKALVYFLIFWVILLSLVINIFASIF YTPLVSYLHQYVPHVVLLEFSGILSIFIKYFLKSLTLMLLFLAVLTPLYFIPFIFGVF GVFFSIVPHFLFFKNTMSLDIASMIFNHQSYQNLLKQHRLKHRYRFFCYFLPSLIPFF NFFATLLQTLMLTHYFFIFKEKEC"
		7361. .8260
		/gene="jhp0161"
		7361. .8260
		/gene="jhp0161"
gene		/note="similar to H. pylori 26695 gene HP0175"
		/codon_start=1
		/transl_table=11
		/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
		/protein_id="AAD05744.1"
CDS		/db_xref="GI:4154677"
		/translation="MKKNILNALVGLSASFLMAKPAHNANNSTHNTKETTDSAGV LATVDGRPTIKQRPDMIKQRPNFDFOKLKEKEKEALTEQAIRTALVNEAKAEKLNQ TPEFKAMWEAVKQALVEFWAKQAEVKKIQIPEKEMQDFYNANKDQLFVKQEAHAR
		complement(4221. .5393)
		/gene="moea"
		/note="synonym: jhp0158"
CDS		complement(4221. .5393)
		/gene="moea"
		/note="similar to H. pylori 26695 gene HP0172"
		/codon_start=1
		/transl_table=11
		/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
gene		/protein_id="AAD05741.1"
		/db_xref="GI:4154674"
		/translation="MISFKEALKITHSNIPKLEPVEVVSLEFESAGRILAEILICVHAL PKFQNSAMDGCFKMDIGQKTQIHI FAGDVSALVENECVKIMTGAMVPKGIE TIVFIECMLESHKDFALAPKDFKIHANIQKGENASNLVLPKNTLNYNTHGIALIAS OGPKEIKAFRLKIALPSSGDELVPAGONALECOVVDYNSGVENMLKNYNTHFTGLV KDDKNLQKILELQGYDVLSSAGVSVDGDFPKDALKERNALFYEKVNLPKGPVT LAQLNQSIILGPGNLSCLLVRLVILFLLESLNLDKFKLPKPAQINAPLKNKK RTHILGNYSNHQFIPYNNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFEN"
		complement(5403. .6170)
		/gene="flir"
CDS		/note="synonym: jhp0159"
		complement(5403. .6170)
		/gene="flir"
		/note="similar to H. pylori 26695 gene HP0173"
		/codon_start=1
gene		/transl_table=11
		/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
		/protein_id="AAD05742.1"
		/db_xref="GI:4154675"
		/translation="MLDPIQELSTPHVRDFLLFLRVSGVLSFPFFFNHVLVPLSVRG ALSLYVAIFYPTLEFNAIYTPGFIACELCFLGVCASVFFQIVFASLVFAFDSI SFSMGLTMASAYDPISGOKPIVGOALLLAILLLDLSFHHOILLFVDHSLKAVPLG RFVPEPELAKNIVKAFSHLVIGFSMAPPIICLVLLSDIIFGMIMKTHPQFNALLAIFG PVKIAIGFVGHILLASAIMGRFKEISLAFSVISKIP"
CDS		complement(6164. .6892)
		/gene="jhp0160"
		complement(6164. .6892)
		/gene="jhp0160"
		/note="similar to H. pylori 26695 gene HP0174"
gene		/codon_start=1
		/transl_table=11
		/product="putative"
		/protein_id="AAD05743.1"
		/db_xref="GI:4154676"
CDS		/translation="MVLFLSIFKSFNDFLSARMLLINLGPILLSLAFSGAIFYVNGG SIVNYCQTLLEQSLNDYAHSGQFFAGVPAWFKALVYFLIFWVILLSLVINIFASIF YTPLVSYLHQYVPHVVLLEFSGILSIFIKYFLKSLTLMLLFLAVLTPLYFIPFIFGVF GVFFSIVPHFLFFKNTMSLDIASMIFNHQSYQNLLKQHRLKHRYRFFCYFLPSLIPFF NFFATLLQTLMLTHYFFIFKEKEC"
		7361. .8260
		/gene="jhp0161"
		7361. .8260
		/gene="jhp0161"
gene		/note="similar to H. pylori 26695 gene HP0175"
		/codon_start=1
		/transl_table=11
		/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
		/protein_id="AAD05744.1"
CDS		/db_xref="GI:4154677"
		/translation="MKKNILNALVGLSASFLMAKPAHNANNSTHNTKETTDSAGV LATVDGRPTIKQRPDMIKQRPNFDFOKLKEKEKEALTEQAIRTALVNEAKAEKLNQ TPEFKAMWEAVKQALVEFWAKQAEVKKIQIPEKEMQDFYNANKDQLFVKQEAHAR
		complement(4221. .5393)
		/gene="moea"
		/note="synonym: jhp0158"
CDS		complement(4221. .5393)
		/gene="moea"
		/note="similar to H. pylori 26695 gene HP0172"
		/codon_start=1
		/transl_table=11
		/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
gene		/protein_id="AAD05741.1"
		/db_xref="GI:4154674"
		/translation="MISFKEALKITHSNIPKLEPVEVVSLEFESAGRILAEILICVHAL PKFQNSAMDGCFKMDIGQKTQIHI FAGDVSALVENECVKIMTGAMVPKGIE TIVFIECMLESHKDFALAPKDFKIHANIQKGENASNLVLPKNTLNYNTHGIALIAS OGPKEIKAFRLKIALPSSGDELVPAGONALECOVVDYNSGVENMLKNYNTHFTGLV KDDKNLQKILELQGYDVLSSAGVSVDGDFPKDALKERNALFYEKVNLPKGPVT LAQLNQSIILGPGNLSCLLVRLVILFLLESLNLDKFKLPKPAQINAPLKNKK RTHILGNYSNHQFIPYNNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFEN"
		complement(5403. .6170)
		/gene="flir"
CDS		/note="synonym: jhp0159"
		complement(5403. .6170)
		/gene="flir"
		/note="similar to H. pylori 26695 gene HP0173"
		/codon_start=1
gene		/transl_table=11
		/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
		/protein_id="AAD05742.1"
		/db_xref="GI:4154675"
		/translation="MLDPIQELSTPHVRDFLLFLRVSGVLSFPFFFNHVLVPLSVRG ALSLYVAIFYPTLEFNAIYTPGFIACELCFLGVCASVFFQIVFASLVFAFDSI SFSMGLTMASAYDPISGOKPIVGOALLLAILLLDLSFHHOILLFVDHSLKAVPLG RFVPEPELAKNIVKAFSHLVIGFSMAPPIICLVLLSDIIFGMIMKTHPQFNALLAIFG PVKIAIGFVGHILLASAIMGRFKEISLAFSVISKIP"
CDS		complement(6164. .6892)
		/gene="jhp0160"
		complement(6164. .6892)
		/gene="jhp0160"
		/note="similar to H. pylori 26695 gene HP0174"
gene		/codon_start=1
		/transl_table=11
		/product="putative"
		/protein_id="AAD05743.1"
		/db_xref="GI:4154676"
CDS		/translation="MVLFLSIFKSFNDFLSARMLLINLGPILLSLAFSGAIFYVNGG SIVNYCQTLLEQSLNDYAHSGQFFAGVPAWFKALVYFLIFWVILLSLVINIFASIF YTPLVSYLHQYVPHVVLLEFSGILSIFIKYFLKSLTLMLLFLAVLTPLYFIPFIFGVF GVFFSIVPHFLFFKNTMSLDIASMIFNHQSYQNLLKQHRLKHRYRFFCYFLPSLIPFF NFFATLLQTLMLTHYFFIFKEKEC"
		7361. .8260
		/gene="jhp0161"
		7361. .8260
		/gene="jhp0161"
gene		/note="similar to H. pylori 26695 gene HP0175"
		/codon_start=1
		/transl_table=11
		/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
		/protein_id="AAD05744.1"
CDS		/db_xref="GI:4154677"
		/translation="MKKNILNALVGLSASFLMAKPAHNANNSTHNTKETTDSAGV LATVDGRPTIKQRPDMIKQRPNFDFOKLKEKEKEALTEQAIRTALVNEAKAEKLNQ TPEFKAMWEAVKQALVEFWAKQAEVKKIQIPEKEMQDFYNANKDQLFVKQEAHAR
		complement(4221. .5393)
		/gene="moea"
		/note="synonym: jhp0158"
CDS		complement(4221. .5393)
		/gene="moea"
		/note="similar to H. pylori 26695 gene HP0172"
		/codon_start=1
		/transl_table=11
		/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
gene		/protein_id="AAD05741.1"
		/db_xref="GI:4154674"
		/translation="MISFKEALKITHSNIPKLEPVEVVSLEFESAGRILAEILICVHAL PKFQNSAMDGCFKMDIGQKTQIHI FAGDVSALVENECVKIMTGAMVPKGIE TIVFIECMLESHKDFALAPKDFKIHANIQKGENASNLVLPKNTLNYNTHGIALIAS OGPKEIKAFRLKIALPSSGDELVPAGONALECOVVDYNSGVENMLKNYNTHFTGLV KDDKNLQKILELQGYDVLSSAGVSVDGDFPKDALKERNALFYEKVNLPKGPVT LAQLNQSIILGPGNLSCLLVRLVILFLLESLNLDKFKLPKPAQINAPLKNKK RTHILGNYSNHQFIPYNNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFEN"
		complement(5403. .6170)
		/gene="flir"
CDS		/note="synonym: jhp0159"
		complement(5403. .6170)
		/gene="flir"
		/note="similar to H. pylori 26695 gene HP0173"
		/codon_start=1
gene		/transl_table=11
		/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
		/protein_id="AAD05742.1"
		/db_xref="GI:4154675"
		/translation="MLDPIQELSTPHVRDFLLFLRVSGVLSFPFFFNHVLVPLSVRG ALSLYVAIFYPTLEFNAIYTPGFIACELCFLGVCASVFFQIVFASLVFAFDSI SFSMGLTMASAYDPISGOKPIVGOALLLAILLLDLSFHHOILLFVDHSLKAVPLG RFVPEPELAKNIVKAFSHLVIGFSMAPPIICLVLLSDIIFGMIMKTHPQFNALLAIFG PVKIAIGFVGHILLASAIMGRFKEISLAFSVISKIP"
CDS		complement(6164. .6892)
		/gene="jhp0160"
		complement(6164. .6892)
		/gene="jhp0160"
		/note="similar to H. pylori 26695 gene HP0174"
gene		/codon_start=1
		/transl_table=11
		/product="putative"
		/protein_id="AAD05743.1"
		/db_xref="GI:4154676"
CDS		/translation="MVLFLSIFKSFNDFLSARMLLINLGPILLSLAFSGAIFYVNGG SIVNYCQTLLEQSLNDYAHSGQFFAGVPAWFKALVYFLIFWVILLSLVINIFASIF YTPLVSYLHQYVPHVVLLEFSGILSIFIKYFLKSLTLMLLFLAVLTPLYFIPFIFGVF GVFFSIVPHFLFFKNTMSLDIASMIFNHQSYQNLLKQHRLKHRYRFFCYFLPSLIPFF NFFATLLQTLMLTHYFFIFKEKEC"
		7361. .8260
		/gene="jhp0161"
		7361. .8260
		/gene="jhp0161"
gene		/note="similar to H. pylori 26695 gene HP0175"
		/codon_start=1
		/transl_table=11
		/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
		/protein_id="AAD05744.1"
CDS		/db_xref="GI:4154677"
		/translation="MKKNILNALVGLSASFLMAKPAHNANNSTHNTKETTDSAGV LATVDGRPTIKQRPDMIKQRPNFDFOKLKEKEKEALTEQAIRTALVNEAKAEKLNQ TPEFKAMWEAVKQALVEFWAKQAEVKKIQIPEKEMQDFYNANKDQLFVKQEAHAR
		complement(4221. .5393)
		/gene="moea"
		/note="synonym: jhp0158"
CDS		complement(4221. .5393)
		/gene="moea"
		/note="similar to H. pylori 26695 gene HP0172"
		/codon_start=1
		/transl_table=11
		/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
gene		/protein_id="AAD05741.1"
		/db_xref="GI:4154674"
		/translation="MISFKEALKITHSNIPKLEPVEVVSLEFESAGRILAEILICVHAL PKFQNSAMDGCFKMDIGQKTQIHI FAGDVSALVENECVKIMTGAMVPKGIE TIVFIECMLESHKDFALAPKDFKIHANIQKGENASNLVLPKNTLNYNTHGIALIAS OGPKEIKAFRLKIALPSSGDELVPAGONALECOVVDYNSGVENMLKNYNTHFTGLV KDDKNLQKILELQGYDVLSSAGVSVDGDFPKDALKERNALFYEKVNLPKGPVT LAQLNQSIILGPGNLSCLLVRLVILFLLESLNLDKFKLPKPAQINAPLKNKK RTHILGNYSNHQFIPYNNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFEN"
		complement(5403. .6170)
		/gene="flir"
CDS		/note="synonym: jhp0159"
		complement(5403. .6170)
		/gene="flir"
		/note="similar to H. pylori 26695 gene HP0173"
		/codon_start=1
gene		/transl_table=11
		/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
		/protein_id="AAD05742.1"
		/db_xref="GI:4154675"
		/translation="MLDPIQELSTPHVRDFLLFLRVSGVLSFPFFFNHVLVPLSVRG ALSLYVAIFYPTLEFNAIYTPGFIACELCFLGVCASVFFQIVFASLVFAFDSI SFSMGLTMASAYDPISGOKPIVGOALLLAILLLDLSFHHOILLFVDHSLKAVPLG RFVPEPELAKNIVKAFSHLVIGFSMAPPIICLVLLSDIIFGMIMKTHPQFNALLAIFG PVKIAIGFVGHILLASAIMGRFKEISLAFSVISKIP"
CDS		complement(6164. .6892)
		/gene="jhp0160"
		complement(6164. .6892)
		/gene="jhp0160"
		/note="similar to H. pylori 26695 gene HP0174"
gene		/codon_start=1
		/transl_table=11
		/product="putative"
		/protein_id="AAD05743.1"
		/db_xref="GI:4154676"
CDS		/translation="MVLFLSIFKSFNDFLSARMLLINLGPILLSLAFSGAIFYVNGG SIVNYCQTLLEQSLNDYAHSGQFFAGVPAWFKALVYFLIFWVILLSLVINIFASIF YTPLVSYLHQYVPHVVLLEFSGILSIFIKYFLKSLTLMLLFLAVLTPLYFIPFIFGVF GVFFSIVPHFLFFKNTMSLDIASMIFNHQSYQNLLKQHRLKHRYRFFCYFLPSLIPFF NFFATLLQTLMLTHYFFIFKEKEC"
		7361. .8260
		/gene="jhp0161"
		7361. .8260
		/gene="jhp0161"
gene		/note="similar to H. pylori 26695 gene HP0175"
		/codon_start=1
		/transl_table=11
		/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
		/protein_id="AAD05744.1"
CDS		/db_xref="GI:4154677"
		/translation="MKKNILNALVGLSASFLMAKPAHNANNSTHNTKETTDSAGV LATVDGRPTIKQRPDMIKQRPNFDFOKLKEKEKEALTEQAIRTALVNEAKAEKLNQ TPEFKAMWEAVKQALVEFWAKQAEVKKIQIPEKEMQDFYNANKDQLFVKQEAHAR
		complement(4221. .5393)
		/gene="moea"
		/note="synonym: jhp0158"
CDS		complement(4221. .5393)
		/gene="moea"
		/note="similar to H. pylori 26695 gene HP0172"
		/codon_start=1
		/transl_table=11
		/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
gene		/protein_id="AAD05741.1"
		/db_xref="GI:4154674"
		/translation="MISFKEALKITHSNIPKLEPVEVVSLEFESAGRILAEILICVHAL PKFQNSAMDGCFKMDIGQKTQIHI FAGDVSALVENECVKIMTGAMVPKGIE TIVFIECMLESHKDFALAPKDFKIHANIQKGENASNLVLPKNTLNYNTHGIALIAS OGPKEIKAFRLKIALPSSGDELVPAGONALECOVVDYNSGVENMLKNYNTHFTGLV KDDKNLQKILELQGYDVLSSAGVSVDGDFPKDALKERNALFYEKVNLPKGPVT LAQLNQSIILGPGNLSCLLVRLVILFLLESLNLDKFKLPKPAQINAPLKNKK RTHILGNYSNHQFIPYNNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFEN"
		complement(5403. .6170)
		/gene="flir"
CDS		/note="synonym: jhp0159"
		complement(5403. .6170)
		/gene="flir"
		/note="similar to H. pylori 26695 gene HP0173"
		/codon_start=1
gene		/transl_table=11
		/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
		/protein_id="AAD05742.1"
		/db_xref="GI:4154675"
		/translation="MLDPIQELSTPHVRDFLLFLRVSGVLSFPFFFNHVLVPLSVRG ALSLYVAIFYPTLEFNAIYTPGFIACELCFLGVCASVFFQIVFASLVFAFDSI SFSMGLTMASAYDPISGOKPIVGOALLLAILLLDLSFHHOILLFVDHSLKAVPLG RFVPEPELAKNIVKAFSHLVIGFSMAPPIICLVLLSDIIFGMIMKTHPQFNALLAIFG PVKIAIGFVGHILLASAIMGRFKEISLAFSVISKIP"
CDS		complement(6164. .6892)
		/gene="jhp0160"
		complement(6164. .6892)
		/gene="jhp0160"
		/note="similar to H. pylori 26695 gene HP0174"
gene		/codon_start=1
		/transl_table=11
		/product="putative"
		/protein_id="AAD05743.1"
		/db_xref="GI:4154676"
CDS		/translation="MVLFLSIFKSFNDFLSARMLLINLGPILLSLAFSGAIFYVNGG SIVNYCQTLLEQSLNDYAHSGQFFAGVPAWFKALVYFLIFWVILLSLVINIFASIF YTPLVSYLHQYVPHVVLLEFSGILSIFIKYFLKSLTLMLLFLAVLTPLYFIPFIFGVF GVFFSIVPHFLFFKNTMSLDIASMIFNHQSYQNLLKQHRLKHRYRFFCYFLPSLIPFF NFFATLLQTLMLTHYFFIFKEKEC"
		7361. .8260
		/gene="jhp0161"
		7361. .8260
		/gene="jhp0161"
gene		/note="similar to H. pylori 26695 gene HP0175"
		/codon_start=1
		/transl_table=11
		/product="putative PEPTIDYL-PROLYL CIS-TRANS ISOMERASE"
		/protein_id="AAD05744.1"
CDS		/db_xref="GI:4154677"
		/translation="MKKNILNALVGLSASFLMAKPAHNANNSTHNTKETTDSAGV LATVDGRPTIKQRPDMIKQRPNFDFOKLKEKEKEALTEQAIRTALVNEAKAEKLNQ TPEFKAMWEAVKQALVEFWAKQAEVKKIQIPEKEMQDFYNANKDQLFVKQEAHAR
		complement(4221. .5393)
		/gene="moea"
		/note="synonym: jhp0158"
CDS		complement(4221. .5393)
		/gene="moea"
		/note="similar to H. pylori 26695 gene HP0172"
		/codon_start=1
		/transl_table=11
		/product="MOLYBDOPTERIN BIOSYNTHESIS PROTEIN"
gene		/protein_id="AAD05741.1"
		/db_xref="GI:4154674"
		/translation="MISFKEALKITHSNIPKLEPVEVVSLEFESAGRILAEILICVHAL PKFQNSAMDGCFKMDIGQKTQIHI FAGDVSALVENECVKIMTGAMVPKGIE TIVFIECMLESHKDFALAPKDFKIHANIQKGENASNLVLPKNTLNYNTHGIALIAS OGPKEIKAFRLKIALPSSGDELVPAGONALECOVVDYNSGVENMLKNYNTHFTGLV KDDKNLQKILELQGYDVLSSAGVSVDGDFPKDALKERNALFYEKVNLPKGPVT LAQLNQSIILGPGNLSCLLVRLVILFLLESLNLDKFKLPKPAQINAPLKNKK RTHILGNYSNHQFIPYNNRYESGAIQALQAVDSIALIDEGVGLVQGSIEILRFEN"
		complement(5403. .6170)
		/gene="flir"
CDS		/note="synonym: jhp0159"
		complement(5403. .6170)
		/gene="flir"
		/note="similar to H. pylori 26695 gene HP0173"
		/codon_start=1
gene		/transl_table=11
		/product="FLAGELLAR BIOSYNTHESIS PROTEIN"
		/protein_id="AAD05742.1"
		/db_xref="GI:4154675"
		/translation="MLDPIQELSTPHVRDFLLFLRVSGVLSFPFFFNHVLVPLSVRG

HLIVKTYDEAKRIISIDIKQPKAKKAFIELANRDTIDPNSKVAQNGDGLKFKQK
 MAFDQSKAALTPGDYTKTPVKTEFYHIIYLIKSDPVTYTYEQAQPTIKGMLOEK
 8274. .9197
 /gene="fba"
 /note="synonym: jhp0162"
 8274. .9197
 /gene="fba"
 /note="similar to H. pylori 26695 gene HP0176"
 /codon start=1
 /transl table=11
 /product="FRUCTOSE-BISPHOSPHATE ALDOLASE"
 /protein_id="AA005734.1"
 /db_xref="GI:1154667"
 /translations="MLVKGNEILLKAKHEGYGVGAFNFVFNELNAIFEAGNEENSP
 FIOASEGAIKYMGIDMAVGMVKIMCERYPHIPVALHLDHGTTFESCEKAVKAGFTSWM

Query Match 100.0%; Score 29; DB 1; Length 10085;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGAACACACAAATGAAAAAATATC 29
 |||||
 DB 7347 GTGAGNACACACANTGAAAAAATATC 7375

RESULT 9
 LOCUS AE000538 11421 bp DNA linear BCT 06-APR-1999
 DEFINITION Helicobacter pylori 26695 section 16 of 134 of the complete genome.
 ACCESSION AE000538 AE000511
 VERSION AE000538.1 GI:2131263
 KEYWORDS
 SOURCE Helicobacter pylori 26695
 ORGANISM Helicobacter pylori 26695
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 Helicobacteraceae; Helicobacter.
 1 (bases 1 to 11421)
 Tomb, J.-F., White, O., Kerslavage, A.R., Clayton, R.A., Sutton, G.G.,
 Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S.,
 Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F.,
 Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G.,
 Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D.,
 Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R.,
 Peterson, J.D., Kelley, J.M., Karp, P.D., Smith, H.O., Fraser, C.M. and
 Venter, J.C.
 The complete genome sequence of the gastric pathogen Helicobacter
 pylori
 Nature 388 (6642), 535-547 (1997)
 JOURNAL
 MEDLINE 97394467
 PUBMED 9252185
 REFERENCE
 2 (bases 1 to 11421)
 Tomb, J.-F., White, O., Kerslavage, A.R., Clayton, R.A., Sutton, G.G.,
 Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S.,
 Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F.,
 Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G.,
 Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D.,
 Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R.,
 Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C.,
 Bowman, C., Watthey, L., Wallin, E., Hayes, W.S., Borodovsky, M.,
 Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.
 Direct Submission
 Submitted (06-AUG-1997) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 3 (bases 1 to 11421)
 White, O.
 Direct Submission
 Submitted (17-MAR-1999) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 JOURNAL
 FEATURES
 source Location/Qualifiers
 1. .11421
 /organism="Helicobacter pylori 26695"
 /mol_type="genomic DNA"

CDS
 /gene="HP0179"
 complement(4205..4846)
 /gene="HP0179"
 /note="similar to GP:1786703 percent identity: 37.62;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="ABC transporter, ATP-binding protein"
 /protein_id="AAD07249.1"
 /db_xref="GI:2313268"
 /translation="MIKAINISHAFKPELYNGVNHIIKPKSLAILGVSGSKSTLLS
 HLATMLKNSGTSISLEHQDIYALNSKLELRILKVGIIIFQSHYLFKGFIALEDPG
 ASILAKINSGSLQLGIAHTLKQGVGELSGQQOQLSIARVLSKKPKIIFIALEDPG
 NLDVTSANOVISMLQNYITEKEGALVLAHDEHFAFTCSQVYVLEKEVLKEK"
 complement(4843..6120)
 /gene="HP0180"
 complement(4843..6120)
 /gene="HP0180"
 /note="similar to GB:L42023 SP:P44626 PID:1003497
 PID:122229 PID:1204555 percent identity: 28.03;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="apolipoprotein N-acyltransferase (cute)"
 /protein_id="AAD07250.1"
 /db_xref="GI:2313269"
 /translation="WRLLFPNCAELLACMFVSVVYVNAVLDAYAIENPYISITLSL
 LAPLMALFKPRNSAFALGFVGFALLFYWCALSFYSDFTYLLPLIIVLIALVIGV
 LFYLLLYPENYFLLSGLSGSIHPFGFMDLPVDSFYSVFRVFKLSGLGLFIACI
 FLSTPLKRIIGVLLGLADFNPKTSDLVKGNIELVSTKPODLKDFSSYND
 IENNLIKETIAQSKQKTLIVPEPTAIENSPFKAKLEDLSNIIAIGTLRTQG
 YNLNSFLFSESQIADKVLPAFGETMPLFLOKPLEKLPFGESTIYLRNAPHF
 SDFTLDDTFRLPICVETSKPAYSNSPKIFIVMSNNWSPSPSEPTLQRTLLKYIA
 RYDKIIILHSANFSTYISLPSLLGDIILFRKKS"
 6378..7052
 /gene="HP0181"
 6378..7052
 /gene="HP0181"
 /note="hypothetical protein; identified by GeneMark;
 putative"
 /codon_start=1
 /transl_table=11
 /product="H. pylori predicted coding region HP0181"
 /protein_id="AAD07253.1"
 /db_xref="GI:2313272"
 /translation="WVVAFGIRGPHGVSEVAGTLGIVLGVYASRYSVAVGNLFS
 EHLYLNETMNLIGFLLVLASIVFFLAFLGVLGKLVLFVSGLGIIDKALGFIPSC
 KTFVLISFLYALSMEYMKDANAYLOEKSAPFTSKVSKIMELDGVHVEQNLKD
 NLEEMSDVKNKESFNKNKESFNKMDKVESLKEKADLPKNMLDPRANQTPPNPTP
 SNKEPL"
 7062..8567
 /gene="HP0182"
 7062..8567
 /gene="HP0182"
 /note="similar to GB:M63448 SP:P41258 PID:144208 percent
 identity: 58.63; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="lysyl-tRNA synthetase (lyss)"
 /protein_id="AAD07251.1"
 /db_xref="GI:2313270"
 /translation="NFSNQYIQORIHKANSLEEGKNPYQNGLKSLTNAAPLEKAY
 VAGLEPKDKCESIVGRKLLRLMGKACFIKVEDESTILQVYVQNELNDEFKSLK
 KHLVGDIVLVKGFPAFTKSGELSIIHALEFHLISKTIIVLPKFKHGLSDIELRYQRY
 LDLIYVPSVKDVKKRSIVSVRFEMEGFLEVEPTMMHPIDPGANARPIITYHNA
 LEVERYLAPLRLIIVGGFEAVFNFRNMGDHSNPPFTWIEFWAYHTY
 EDLIELSKRLFDYLLKTLNDLSKIYNDMEVDFNQTSVSYLDALETIGGISKDIIEK
 EDRLAYLLEQGIKVEPNLYGKLLAEAFDPHVEHQLINPTFTQYPIEISPLARNR
 SNPNIAFRKFLFIAGKFTANGFSELDPLDQLERFNQKQVKEKGEDEQAQWEDYIW
 ALAHGMPTAGGGIGIDRLVMLLTGAKSIKDVILFPAMRPVKNDNFVSESE"
 8567..9817
 /gene="HP0183"

gene
 CDS

CDS
 8567..9817
 /gene="HP0183"
 /note="similar to GB:J01620 SP:P00477 GB:V00283 PID:146218
 PID:41603 percent identity: 53.96; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="serine hydroxymethyltransferase (glyA)"
 /protein_id="AAD07252.1"
 Query Match 100.0%; Score 29; DB 1; Length 11421;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTGGAGAACACACATGAAAAAATATC 29
 |||||
 Db 232 GTGGAGAACACACATGAAAAAATATC 260
 |||||
 RESULT 10
 BD193947/c 4134 bp DNA linear PAT 17-JUL-2003
 LOCUS Enterococcus faecalis polynucleotides and polypeptides.
 DEFINITION BD193947
 ACCESSION BD193947
 VERSION BD193947.1 GI:33003686
 KEYWORDS JP 2002529046-A/541.
 SOURCE unidentified
 ORGANISM unidentified
 unclassified.
 REFERENCE 1 (bases 1 to 4134)
 AUTHORS Kunsch,C.A., Dillon,P.J. and Barash,S.C.
 TITLE Enterococcus faecialis polynucleotides and polypeptides
 JOURNAL Patent: JP 2002529046-A 541 03-SEP-2002;
 HUMAN GENOME SCIENCES INC
 COMMENT OS Unidentified
 PN JP 2002529046-A/541
 PD 03-SEP-2002
 PF 04-MAY-1998 JP 1998548302
 PR 06-MAY-1997 US 60/044031.16-MAY-1997 US 60/046655 PR
 14-NOV-1997 US 60/086009
 PI CHARLES A KUNSCH,PATRICK J DILLON,STEVEN C BARASH PC
 C12N15/31,C07K14/315,C07K16/12,C12Q1/68
 CC Strandedness: Double;
 CC Topology: Linear;
 CC Enterococcus faecialis polynucleotides and polypeptides FH
 Key Location/Qualifiers
 FT source 1..4134
 FEATURES
 source Location/Qualifiers
 1..4134
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 ORIGIN
 Query Match 80.0%; Score 23.2; DB 6; Length 4134;
 Best Local Similarity 89.3%; Pred. No. 2.2e+02;
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TGGAGAACACACATGAAAAAATATC 29
 |||||
 Db 1851 TGGAGAACACACATGAAAAAATATC 1824
 |||||
 RESULT 11
 AE016956/c 304454 bp DNA linear BCT 28-MAR-2003
 LOCUS Enterococcus faecalis V583, section 10 of 11 of the complete
 DEFINITION genome.
 ACCESSION AE016956 AE016830
 VERSION AE016956.1 GI:29344791
 KEYWORDS
 SOURCE Enterococcus faecalis V583

repeat_region	/note="MIR repeat: matches 3. .232 of consensus" 7253. .8047 /note="L1MA4A repeat: matches 5507. .6299 of consensus" 8883. .8981 /note="MERSA repeat: matches 13. .115 of consensus" 9041. .9145 /note="MERSB repeat: matches 1. .110 of consensus" 9589. .9874 /note="AluJo repeat: matches 1. .281 of consensus" 9883. .9940 /note="MERSA repeat: matches 132. .189 of consensus" 10342. .1052 /note="Alusk repeat: matches 1. .296 of consensus" 11405. .11454 /note="25 copies 2 mer aa 78% conserved" 11553. .11845 /note="Alusk repeat: matches 1. .294 of consensus" 12836. .13671 /note="L1MB3A repeat: matches 5237. .6095 of consensus" 13734. .14362 /note="L1M4 repeat: matches 4342. .5025 of consensus" 14363. .14660 /note="AluJb repeat: matches 1. .298 of consensus" 14661. .14801 /note="L1M4 repeat: matches 4211. .4342 of consensus" 15191. .15485 /note="Alusk repeat: matches 1. .294 of consensus" 15495. .15781 /note="Alusk repeat: matches 1. .287 of consensus" 15782. .16353 /note="L1M4 repeat: matches 4911. .5495 of consensus" 16436. .16546 /note="L2 repeat: matches 1795. .1884 of consensus" 17490. .17651 /note="L2 repeat: matches 2600. .2749 of consensus" 17652. .17905 /note="Alusg repeat: matches 1. .253 of consensus" 17906. .18099 /note="L2 repeat: matches 2297. .2600 of consensus" 19226. .19508 /note="AluJb repeat: matches 1. .288 of consensus" 19618. .19826 /note="L2 repeat: matches 2480. .2714 of consensus" 20588. .20764 /note="MIR repeat: matches 6. .190 of consensus" 20970. .21131 /note="L1MB1 repeat: matches 5887. .6042 of consensus" 21132. .21426 /note="Alusk repeat: matches 1. .295 of consensus" 21427. .22078 /note="L1MB1 repeat: matches 5200. .5887 of consensus" 23672. .23963 /note="Alusk repeat: matches 1. .293 of consensus" 24857. .24884 /note="14 copies 2 mer ct 92% conserved" 24886. .25173 /note="AluJb repeat: matches 1. .285 of consensus" 25264. .25395 /note="L1P3 repeat: matches 5606. .5747 of consensus" 25396. .25559 /note="MIR repeat: matches 77. .241 of consensus" 26182. .26479 /note="Alusk repeat: matches 1. .296 of consensus" 26788. .26838 /note="L2 repeat: matches 2660. .2710 of consensus" 26935. .27259 /note="L1MC/D repeat: matches 5347. .5680 of consensus" 29200. .29382 /note="MERSA repeat: matches 2. .189 of consensus" 29450. .29803 /note="L1MA9 repeat: matches 5957. .6306 of consensus" 29949. .30243 /note="Alusg repeat: matches 1. .310 of consensus"
---------------	---

repeat_region	30333. .30460 /note="MERG3A repeat: matches 78. .209 of consensus"
repeat_region	31898. .32190 /note="AluCb repeat: matches 1. .295 of consensus"
repeat_region	33755. .34056 /note="AluCb repeat: matches 1. .311 of consensus"
repeat_region	34637. .34708 /note="Alu repeat: matches 244. .302 of consensus"
repeat_region	34882. .35177 /note="AluCb repeat: matches 1. .306 of consensus"
repeat_region	35496. .35679 /note="MIR repeat: matches 11. .197 of consensus"
repeat_region	35731. .35765 /note="U2 repeat: matches 1. .35 of consensus"
repeat_region	35880. .36143 /note="L2 repeat: matches 2161. .2454 of consensus"
repeat_region	37715. .37752 /note="19 copies 2 mer tt 78% conserved"
repeat_region	39471. .39782 /note="AluCb repeat: matches 1. .309 of consensus"
repeat_region	40626. .40696 /note="MIR repeat: matches 82. .152 of consensus"
misc_feature	41366. .41921 /note="CpG island" /evidence=not_experimental
repeat_region	41835. .41894 /note="35 copies 2 mer gg 65% conserved"
repeat_region	43257. .43322 /note="MIR repeat: matches 82. .145 of consensus"
repeat_region	43974. .44086 /note="MIR repeat: matches 35. .145 of consensus"
repeat_region	44105. .44415 /note="AluCb repeat: matches 1. .310 of consensus"
misc_feature	44140. .45342 /note="CpG island" /evidence=not_experimental
repeat_region	45531. .45784 /note="AluCb repeat: matches 30. .278 of consensus"
repeat_region	46311. .46432 /note="MIR repeat: matches 35. .164 of consensus"

Query Match	77.9%	Score 22.6;	DB 9;	Length 110098;
Best Local Similarity	86.2%	Pred. No. 2.8e+02;		
Matches 25; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY 1 GTGAGAACACACACAATGAATAAATATC 29
|| || || || || || || || || || || ||
Dδ 95371 GTTGAGAACACACACTGAATAAAATTTC 95343

```

RESULT 13
BX088597          linear      VRT 16-MAR-2004
LOCUS             Zebrafish DNA sequence from clone DKEY-171P23 in linkage group 5,
DEFINITION        complete sequence.
ACCESSION         BX088597
VERSION           BX088597.17 GI:45433485
KEYWORDS          HTG.
SOURCE            Danio rerio (zebrafish)
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                  Cypriniformes; Cyprinidae; Danio.
REFERENCE         1 (bases 1 to 171692)
AUTHORS           Ellwood M.
TITLE             Direct Submission
JOURNAL           Submitted (13-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
                  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                  zfiah-halperanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT           On Mar 13, 2004 this sequence version replaced gi:45379346.
                  ----- Genome Center
                  Center: Wellcome Trust Sanger Institute
                  Center code: SC

```

Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhihong Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-171P23 is from a Zebrafish BAC library

VECTOR: pindigobAC-5.

FEATURES

Location/Qualifiers
 1..171692
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="DKEY-171P23"
 /clone_lib="DanioKey"

ORIGIN

Query Match 77.9%; Score 22.6; DB 5; Length 171692;
 Best Local Similarity 86.2%; Pred. No. 2.7e+02;
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACATGAAAAAATATC 29

Db 28427 GTGGAAACATTCATGTAAAAAATATC 28455

RESULT 14

AL133351 173670 bp DNA linear PRI 16-DEC-2001
 LOCUS Human DNA sequence from clone RP1-90J20 on chromosome 6p24.1-25.3,
 DEFINITION complete sequence.

ACCESSION AL133351

VERSION AL133351.34 GI:17907166

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 173670)

Johnston, C.

Submitted (16-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Dec 18, 2001 this sequence version replaced gi:14041741.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>

RP1-90J20 is from the library RP1-1 constructed by the group of

Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone

RP1-90J20 It may be shorter because we sequence overlapping

sections only once, except for a short overlap.

The true left end of clone RP1-90J20 is at 1 in this sequence. The

true left end of clone RP1-40516 is at 173571 in this sequence. The

true right end of clone RP11-42056 is at 6201 in this sequence.

Location/Qualifiers

1..173670
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="p24.1-25.3"
 /clone="RP1-90J20"
 /clone_lib="RP1-1"

ORIGIN

Query Match 77.9%; Score 22.6; DB 9; Length 173670;
 Best Local Similarity 86.2%; Pred. No. 2.7e+02;
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACATGAAAAAATATC 29

Db 88203 GAGGAGAACACACATGAAAAAATATTC 88231

RESULT 15

AC013339/C

LOCUS Homo sapiens clone RP11-20L11, WORKING DRAFT SEQUENCE, 17 unordered
 DEFINITION pieces.

ACCESSION AC013339

VERSION AC013339.4 GI:7341953

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 189058)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens, clone RP11-20L11

Unpublished

2 (bases 1 to 189058)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckhalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Forrester, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lhoczkzy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 30, 2000 this sequence version replaced gi:6573892.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RN/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L3923
 Center clone name: 20_L_11
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 181522 bases at least Q40
 Consensus quality: 184497 bases at least Q30
 Consensus quality: 185786 bases at least Q20
 Insert size: 183000; agarose-efp
 Insert size: 187458; sum-of-contigs
 Quality coverage: 6.2 in Q20 bases; agarose-efp
 Quality coverage: 6.0 in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently
 consists of 17 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1935: contig of 1935 bp in length
 * 1936 2035: gap of 100 bp
 * 2036 3361: contig of 1326 bp in length
 * 3362 3461: gap of 100 bp
 * 3462 5565: contig of 2104 bp in length
 * 5566 5665: gap of 100 bp
 * 5666 8885: contig of 3220 bp in length
 * 8886 13422: contig of 4437 bp in length
 * 13423 13522: gap of 100 bp
 * 13523 16889: contig of 3367 bp in length
 * 16890 16990: gap of 100 bp
 * 16990 23707: contig of 6717 bp in length
 * 23707 23806: gap of 100 bp
 * 23807 29202: contig of 5396 bp in length
 * 29202 29302: gap of 100 bp
 * 29303 40541: contig of 11239 bp in length
 * 40542 40641: gap of 100 bp
 * 40642 49558: contig of 8917 bp in length
 * 49559 49658: gap of 100 bp
 * 49659 60489: contig of 10831 bp in length
 * 60490 60589: gap of 100 bp
 * 60590 71415: contig of 10826 bp in length
 * 71416 71515: gap of 100 bp
 * 71516 85600: contig of 14085 bp in length
 * 85601 85700: gap of 100 bp
 * 85701 102471: contig of 16771 bp in length
 * 102472 102571: gap of 100 bp
 * 102572 123889: contig of 21318 bp in length
 * 123890 123989: gap of 100 bp
 * 123990 148414: contig of 24425 bp in length

FEATURES

source

* 148415 148514: gap of 100 bp
 * 148515 189058: contig of 40544 bp in length.
 Location/Qualifiers
 1..189058

/organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-20L11"
 /clone_lib="RPC1-11 Human Male BAC"

1..1935
 /note="assembly_fragment"
 2036..3361
 /note="assembly_fragment"
 3462..5565
 /note="assembly_fragment"
 5666..8885
 /note="assembly_fragment"
 8986..13422
 /note="assembly_fragment"
 13523..16889
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 16990..23706
 /note="assembly_fragment"
 23807..29202
 /note="assembly_fragment"
 29303..40541
 /note="assembly_fragment"
 40642..49558
 /note="assembly_fragment"
 49659..60489
 /note="assembly_fragment"
 60590..71415
 /note="assembly_fragment"
 71516..85600
 /note="assembly_fragment"
 85701..102471
 /note="assembly_fragment"
 102572..123889
 /note="assembly_fragment"
 clone_end:T7
 vector_side:left
 123990..148414
 /note="assembly_fragment"
 148515..189058
 /note="assembly_fragment"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 77.9%; Score 22.6; DB 2; Length 189058;
 Best Local Similarity 86.2%; Pred. No. 2.7e+02;
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTGGAGAACACACATGAAAAAATATC 29

DB 14486 GAGGAGAACACACATGAAAAAATATTC 14458

Search completed: April 24, 2005, 03:21:42
 Job time : 763.775 secs

THIS PAGE BLANK (USPTO)